Dissertation

Titel der Dissertation

’Towards automated NMR structure determination’

Verfasser

Andreas Schedlbauer

Angestrebter akademischer Grad

Dr. rer. nat.

Wien, 2008

Studienkennzahl lt. Studienblatt:A091419

Dissertationsgebiet lt. Studienblatt:

Betreuerin / Betreuer:Prof. Dr. Robert Konrat

Abstract

ICln (protein associated with a nucleotide sensitive chlorid current; I = current, Cl^- = chloride, n = nucleotide) initially cloned from Madin Darby canine kidney (MDCK) cells is a 235 residues (27 kDa) ubiquitously expressed abundant protein and highly conserved among distant species. It is essential for cell survival and involved in different critical cellular pathways as diverse as cell volume homeostasis, angiogenesis, or RNA processing. From lipid bilayer reconstitution experiments and some other data ICln was originally published to to be the swelling-induced nucleotide sensitive chloride channel itself by forming a homo-dimer, at which the trans-membrane pore is establised by an eight-stranded, anti-parallel β-barrel. However, although specific knock down experiments indicated a direct link between ICln and regulated cell volume degrease, closer comparison of the biophysical characteristics of the Cl^- currents elicited in cultured cells under hypo-osmotic conditions, $I_{Cl,swell}$ (found to be remarkable constant among different cell types), with those of ICln when overexpressed in *Xenopus oocytes* revealed some differences. Furthermore, immunofluorenscence studies showed a primarily localization in the cytoplasm under normal condition rather than next to the plasma membrane (at which the protein is translocated under hypo-osmotic stress). These findings led to some doubt on a direct channel function of ICln and rather suggest a regulatory role of the protein. The Identification of the $I_{Cl,swell}$ related chloride channels and their regulatory proteins is currently the focus of extensive research because changes in the chloride permeability of cells are the basics of various human inherited diseases, as Dent's disease, osteopetrosis, and Bartter syndrome.

To provide structural information on the molecular level the tertiary conformation of *Canis familiaris* ICln in solution was determined in this work. The N-terminal part of ICln was found to fold in a Pleckstrine homology-like structure (a module typically found in proteins involved in cell signaling), whereas its C-terminal region was examined to be intrinsic unfoled with some low extent secondary structure preferences. In addition, from titration experiments ICln was shown to interact with the membrane-proximal regions of the platelet $\alpha_{2\beta}$ integrin protein (M987-WKVGFFKRN-R997) and to LSm4, a protein involved in snRNP biogenisis. In contrast to other PH domain solved per date ICln lacks an electro-statically polarization and is unable to bind phosphoinositols. Finally, the structural and dynamical data obtained provide not any support for a channel function of ICln.

Another analysis was addressed with structure determination of the *homo sapiens* 168 residue

protein Cyclophilin D, a member of the immunophilin familiy which is part of the mitochondrial permeability (MP) transition pore. Small molecular weight inhibitors of Cyclophilin D allowing brain penetration give raise to suppression of the MP transition (as part of neuro apotosis) and results in neuro protection in animal models, thus faciliating a interesting drug target in neurodegenerative diseases. The molecular details of the solution structure and a comparison with the subsequent published x-ray structure of the K133I mutant were given.

Furthermore, an already analyzed structure of the LIM1 domain of the $quail$ CRP2 protein (adopting a histidine-cysteine double zinc finger motif), was refined in order to provied an improved model for subsequent diffusion tensor analysis. By dint of an double uniformly ^{13}C,^{15}N-labeled sample (containing the first 83 residue of CRP2) a fourfold number of long range NOE contacts could be infered resulting in a significantly increased structural precision.

Another topic of this work was the structural characterization of intrinsic unfolded proteins. The secondary structure preferencies of *gallus gallus* brain abundant signal protein 1 (BASP1) and *coturnix coturnix japonica* osteopontin (OPN), both promising candidate targets for cancer therapy, were evaluated. The more pronounced chemical shift dispersion of OPN facilitates the measurement of several dynamic parameters and subsequent dynamical description of the amino acid sequence employing spectral density mapping.

According to the massive efforts of structural genomics an automated 'direct method' approach employing residue type specific labeling for spectra simplification was derived and tested which circumvent the most time consuming steps (sequential chemical shift and NOESY assignment) in NMR structure determination.

The underlying concept of this approach use the fact that NOESY spectra of a protein typically provide a large amount of distance information. Due to the complete absence of any sequential assignment the initially obtained structural results are represented by individual atomic densities ('clouds') labeled by their unassigned chemical shifts, their NOE contacts and other NMR data, if available.

After the density related interatomic distances are employed for model-driven backbone assignment (in context with selective labeling schema) a covalent model can be build into that structures in an analogical manner to X-ray crystallography, where the structural model is constructed into an electron density map. However, to obtain a sufficient converged (spatial resolved) final atomic distribution essential for that task the procedure iteratively derive new distance restraints from statistical distance and input NOE density analysis to get along with a minimal number of unambiguous experimental input NOE distances.

The feasibility of the protocol was evaluated using a test set of proteins of different size, fold topology, chemical shift dispersion and quality of the input triple resonance data used to derive a covalent ensemble. Additionally, the raw unassigned 'cloud' structure can be applied e.g. in combination with database mining and molecular fragment replacement, or, if a homology model is available, for rapid fold validation.

At last, the capability of an in house made automated binding site mapping schema using unassigned triple resonance spectra and 3D coordinates of the *apo* structure only was verfied using an test set of different fold topologies, ligand sizes, and extents of available chemical shifts.

Zusammenfassung

ICln ist ein hochkonserviertes, in allen eukaryontischen Zellen vorkommendes 27kDa protein und wurde erstmals aus Madin Darby canine kidney (MDCK) Zellen kloniert. Es ist für die vitale Zellfunktion unbedingt erforderlich und involviert in verschiedenen Signaltransduktionswegen wie Zellvolumsregulation, Angiogenese, oder RNA Prozessierung.

Basierend auf Rekonstitutionsexperimente in Membranlipiden und anderen Ergebnissen wurde eine chloridkanalbildende Funktion von ICln vorgeschlagen als Antwort auf hypotonischen Zellstreß, wobei als porengenerierende Einheit eine acht β-Faltblatt umfassendes anti-paralleles β-Faßstruktur gebildet aus einem Homo-Dimer publiziert wurde. Obwohl spezifische knockdown Experimente eine direkte funktionelle Eigenschaft dieses Proteins in der Zellvolumsregulation anzeigten, wurde durch näheren Vergleich der biophysikalischen Charakteristika von Chloridströmen aus Zellen unter hypoosmotischen Bedingungen, $I_{Cl,swell}$ (die in allen bislang untersuchten Zelltypen praktisch gleich sind), mit jenen erhalten aus ICln-überexpression in *Xenopus oocyten* einige deutliche Unterschiede sichtbar. Weiters zeigten Immunofluorenszenzexperimente unter normalen Bedingungen eine primäre zytoplasmatische Lokalisation des Proteins an, das erst durch hypoosmotische Induktion in oder nahe zur Plasmamembran translokalisiert wird. Diese Ergebnisse ließen einige Zweifel an einer porenbildenden Kanalfunktion von ICln aufkommen und viele Forschern ordneten dem Protein in diesem Zusammenhang daraufhin eine rein regulatorische Rolle zu. Nichtsdestoweniger ist die Identifizierung oder der Ausschlußpotentieller chloridkanalbildender Kanditaten enorm wichtig, da physiologische Veränderungen in der Chloridepermeabilität die Grundlage vieler schwerer Erkrankungen, wie der Osteopetrosis, der Dent'sche Krankheit oder des Bartter Syndroms sind.

Zur Generierung konformeller Information auf der molekularen Ebbene war die Strukturbestimmung von ICln in Lösung Teil dieser Arbeit. Dabei zeigte sich das der N-terminale Teil des Proteins in eine Pleckstrin Homologiedomän-analoge Struktur faltet (ein Strukturmotif das bereits in vielen signaltranduktionsregulatorischen Protein identifiziert wurde), wohingegen der C-terminale Abschnitt intrinisch unfaltet ist mit einigen wenigen Sequenzbereichen mit schwach ausgeprägter Sekundärstrukturpreferenz. Weiters konnte durch Titrationsexperimente eine Bindung mit dem membrannahen cytoplasmatischen Teil des Blutplätchen $\alpha_{2\beta}$ Integrinprotein (M987-WKVGFFKRN-R997) sowie auch eine Wechselwirkung mit dem Faktor LSm4, einem funktionellen Protein in der snRNP Biogenese nachgewiesen werden. Im Gegensatz zu anderen bisher gelösten PH-Domän Proteinen bindet ICln nicht an Phosphoinositolen und zeigt

auch keine strukturelle elektrostatische Polarisation. Letztendlich konnte durch keine der erhaltenen dynamischen oder strukturellen NMR-Daten auf eine Kanalfunktion des Proteins geschlossen werden.

Eine weitere Strukturanalyse befaßte sich mit der Bestimmung der Lösungskonformation von Cyclophilin D, einem Mitglied der Immunophilin Familie, das an den mitochondrialen Permeabilitätsporenapparat bindet. Es konnte gezeigt werden, daßniedermolekulare Bindingspartner von Cyclophilin D (die wie cyclosporin A einfach die Blut-Hirn Schranke überwinden können) die Aktivierung dieses Transmembranporenkomplexes und damit die neuronale Apotose verhindern können. Ein Umstand der in Tiermodellen zur Untersuchung von neurodegenerativen Erkrannkungen zum Erhalt der Nervenzellen führte. Aus diesen Gründen stellt Cyclophilin D in diesem Zusammenhang ein interessantes Zielprotein dar. Neben der Tertiärstruktur des Proteins ist auch ein Vergleich mit einer kurz darauf veröffentlicheter Röntgenstruktur einer K133I Punktmutante angegeben.

Weiters wurde eine bereits publizierte NMR-Struktur der LIM1 Domäne aus CRP2 verfeinert. Die Domäne faltet in ein Doppelzinkfingermotif und nimmt bei einer Reihe von Proteininteraktionen eine Adapterfunktion ein. Durch Einsatz einer ^{13}C,^{15}N-doppelt markierten Probe (beinhaltend die ersten 83 Aminosäuren der CRP2 Sequenz) konnte eine vierfache Anzahl an 'long-range' NOE Kontakten für die Strukturbestimmung abgeleitet werden, was zu einer significanten Verbesserung der Strukturmodellpräzision führte.

Ein anderer Teil dieser Arbeit beschäftigte sich mit der strukturellen Charakterisierung von intrinisch ungefaltenen Proteinen. Die Preferenz für Sekundärstrukturelemente von *gallus gallus* BASP1 and *coturnix coturnix japonica* Osteopontin (OPN), beide sehr vielversprechende Zielproteine für die Krebsforschung wurde bestimmt. Die deutlich bessere Dispersion der chemischen Verschiebung für OPN erlaubte weiters die Ableitung dynamischer Meßgrößen und darauffolgender Beschreibung der Moleküldynamik durch 'mapping' der spektralen Dichte.

In Anlehnung zur Bewältigung des enormen Aufwands eines 'structural genomic' Projekts wurde eine einfache sogenannte 'direkte' Methode zur raschen Strukturcharakterisierung in dieser Arbeit entwickelt und anhand experimenteller Daten getestet. Im ersten Schritt wird direkt eine dreidimensionale Anordnung von Atomen berechnet ohne zuvor höchst zeitaufwendig die zugehörigen Spinsysteme samt NOEs zuzuordnen. Dabei wird der Umstand ausgenützt, das NOESY Spektren eines Proteins typischerweise eine sehr große Menge an Distanzinformation (in Form von nichtzugeordnetnen individuellen NOE Kontakten) enthalten. Als strukturelles Ergebnis erhaltet man zunächst durch das Fehlen der sequentiellen Zuordnung (anstelle von

Strukturensembles mit kovalenten Bindungen) durch individuelle NMR-Meßgrößen (chemische Verschiebung, skalare Kopplungen, NOE Muster) markierte Atomdichten des Proteins. Anschließend dienen die erhaltenen interatomaren Abstände als Eingangsinformation für eine simulated-annealing Monte-Carlo ('model-driven') Backbone Zuordnung, die eine kovalente Verküpfung der einzelnen Atomdichten ermöglicht analog der strukturellen Modellkonstruktion durch die Elektronendichte in der Röntgenkristallographie. Um für die Generierung gutaufgelöster Atomdichten mit einer minimalen Anzahl nichtzugeordneter jedoch eindeutig definierter experimenteller Eingangs-NOE-Kontakten das Auskommen zu finden, bildet das Programm in aufeinander folgenden Zyklen der Strukturrechnung durch statistische Analyse der resultierenden interatomaren Abstände und der zugrundeliegenden experimentellen NOE Dichte neue Distanzeinschränkungen für die nächste Runde.

Die Durchführbarkeit der Routine wurde anhang der Testreihen von Proteinen mit unterschiedlicher Molekülgröße, Strukturtopologie, Dispersion der chemischen Verschiebung und Qualität der triple-Resonanz Eingangsdaten für die sequentielle Verküpfung validiert.

Zuletzt wurde das Leistungsvermögen eines von unserem Gruppenleiter verfaßtes Programm für die automatische Vorhersage der Ligandbindungstelle anhand verschiedener Protein-Ligand Komplexe aus der Datenbank getestet.

Contents

1 Introduction **12**

1.1 Biological context of proteins determined by conventional NMR structure analysis 12

1.1.1 ICln . . . 12

1.1.2 Cyclophilin D . . . 20

1.1.3 LIM1 domain from quail cysteine- and glycine-rich protein CRP2 . . . 24

1.1.4 Brain abundant signal protein 1 - BASP1 . . . 26

1.1.5 Osteopontin . . . 28

1.2 Towards automated fold generation . . . 33

1.2.1 Algorithms for automated NOESY assignment demanding prior chemical shift assignment . . . 35

1.2.2 Structure calculation without a priory knowledge of chemical shifts . . . 39

1.2.3 Motivation . . . 43

2 Methods **44**

2.1 Protein NMR spectroscopy . . . 44

2.1.1 Introduction . . . 44

2.1.2 NMR tools and building blocks . . . 45

2.1.3 Heteronuclear experiments for chemical shift assignment . . . 50

2.1.4 Measurement of three bond couplings . . . 57

2.1.5 NOESY experiments . . . 58

2.1.6 Relaxation experiments . . . 59

2.1.7 Processing . . . 61

2.1.8 Resonance assignment strategy . . . 65

2.1.9 Ligand binding in NMR . . . 70

2.1.10 Relaxation, molecular motion and spectral density mapping . . . 72

2.2 NMR and empirical derived structural restraints 74
2.2.1 NOE distance restraints 74
2.2.2 Dihedral angle restraints 76
2.2.3 Conformational database potential 79
2.2.4 Radius of gyration for improving the model packing 80
2.3 Structure determination of proteins 81
2.3.1 Molecular dynamic simulated annealing 81
2.4 Quality of NMR structures 85
2.4.1 Precision of the structure bundle 85
2.4.2 Number and quality of experimental restraints 86
2.4.3 Deviation from idealized covalent geometry 86
2.4.4 Ramachandran maps 86
2.4.5 Contact and environment analysis 89
2.5 Towards automated structure generation 96
2.5.1 Automated backbone assignment using the MONTE program 96
2.5.2 Simplification and sensitivity enhancement of NMR spectra using protein deuteration and specific labeling patterns 100
2.5.3 Protocol for automated methyl group assignment 104
2.6 Semi-automated protein structure calculation with a priori sequential backbone resonance assignment 107
2.7 A direct method protocol employing residue type specific isotope labeling for rapid structure validation and model-driven sequential assignment 108
2.7.1 Input Data and Systems Studied 110
2.7.2 Restraint Classification 112
2.7.3 Structure Calculations: Simulated Annealing I 113
2.7.4 Statistical Analysis of Conserved Distances 113
2.7.5 Simulated Annealing II and Statistical Analysis II 114
2.7.6 Model Driven Sequential Chemical Shift Assignment 117
2.7.7 Simulated annealing III 117
2.8 Automated Protein-Ligand Binding Site Mapping Using Unassigned Triple-Resonance Data 118

3 Experimental **121**

3.1 Samples and NMR parameters . . . 121

3.1.1 ICln . . . 121

3.1.2 Cyclophilin D . . . 125

3.1.3 LIM1 of CRP2 . . . 125

3.1.4 BASP1 . . . 125

3.1.5 Osteopontin . . . 126

3.1.6 Materials . . . 126

3.1.7 Experimental NMR parameter . . . 127

3.2 NMR derived structural restraints . . . 137

3.2.1 Distance restraints . . . 137

3.2.2 Hydrogen bonds . . . 137

3.2.3 Dihedral angle restraints . . . 137

3.3 Protein dynamic NMR measurements . . . 139

3.3.1 Backbone ^{15}N relaxation of ICln . . . 139

3.3.2 Backbone ^{15}N relaxation measurements and spectral density analysis for BASP1 and Osteopontin . . . 140

3.4 NMR structure determination . . . 140

3.4.1 The forcefield . . . 141

3.4.2 Generating the structure templates . . . 141

3.4.3 Molecular dynamic simulating annealing protocol . . . 142

3.4.4 Validation of the obtained structure models . . . 143

3.5 Towards automated protein fold generation . . . 144

3.5.1 Semi-automated protein structure calculation with a priori sequential backbone resonance assignment . . . 144

3.5.2 Rapid protein fold validation using a direct method in context with residue type specific isotope labeling . . . 146

3.6 Automated binding site mapping using ROSA . . . 150

4 Results and Discussion **153**

4.1 NMR analysis of folded proteins . . . 153

4.1.1 ICln . . . 153

4.1.2 Cyclophilin D . . . 180

4.1.3 LIM1 195
4.2 Resonance assignment and relaxation of intrinsic unfolded proteins 203
4.2.1 BASP1 203
4.2.2 Osteopontin 209
4.3 Automated chemical shift assignment of methyl groups 220
4.4 Semi-automated protein structure calculation with a priori sequential backbone resonance assignment 222
4.5 Rapid protein fold validation using a direct method in context with residue type specific isotope labeling 228
4.5.1 Extraction of articial intramolecular distance restraints 228
4.5.2 Practical considerations and structural results 233
4.6 Capability of the Automated binding site mapping schedule (ROSA) 251

Chapter 1

Introduction

1.1 Biological context of proteins determined by conventional NMR structure analysis

1.1.1 ICln

ICln (protein associated with a nucleotide sensitive chloride current; I = current, Cl^- = chloride, n = nucleotide) initially cloned from Madin Darby canine kidney (MDCK) cells is a small (27 kDa) ubiquitously expressed abundant protein and highly conserved among distant species (see figure 1.1). It offers a high amino acid sequence homology ($>$99%) within all human cell lines investigated, and human ICln display 90% and 92% homology to rat and MDCK ICln, respectively [67]. It is essential for cell survival. Functional knock out of ICln in mouse or nematode was found to be lethal in the very early stages of development [142]. ICln was shown to be phosphorylated constitutively primarily on serin residues [156]. The phosphorylation sites of *dog* ICln as verified by mass spectroscopy (Fuerst and co-workers, unpublished) are shown in figure 1.1.

It is a soluble protein under normal condition primarily present in the cytosol [88], where it interact directly with several core spliceosomal (SM) and like SM (LSM) proteins, with the arginine methyl transferase PRTMT5 (also denoted as Janus kinase binding protein 1) [42], all involved in RNA splicing processes.

ICln was reported to bind a yeast Skb1 homologue protein. Skb1 is known to interact with Shk1, which in turn is a homologue of the P21 activated protein kinase (PAK). Because of the known involvement of PAKs in cytoskeleton rearrangement ICln may be linked to a cell morphology regulating system.

Figure 1.1: Multiple alignment of ICln sequences: Canis familiaris (dog), Homo sapiens (human), Mus musculus (mouse), Oryctolagus cuniculus (rabbit), Rattus norvegicus (rat), Xenopus laevis (african clawed frog). Columns with complete identical residue types are boxed in blue, whereas columns are framed in lightred, if more than 70% of its (blue colored) residues are considered as related according Risler similarity scores matrices [149]). Alignment was performed using kAlgin [97], for visualization the program ESPript version 2.2 [49] and Coreldraw was employed. Arrows in yellow indicate the phosphorylation sites of *dog* ICln derived from mass spectroscopy. The three acid domians of ICln are denoted as AD1 to AD3.

Furthermore, ICln binds to cytoskeletal proteins such as β-actin, erythrocyte protein 4.1, or the non muscle isoform of the alkali myosin light chain [38].

Under hypo-tonic cell conditions the main fraction of ICln is found located next to the cell membrane. Heterologous expression of ICln in Xenopus lavis oocytes led to the induction of a chloride conductance closely resembling those elicited by swelling of mammalian cells [130]. Therefore, a role for ICln in cell volume regulation (CVR) was proposed.

The involvement of ICln in different critical cellular pathways, as diverse as RNA processing and CVR, affirmed the pivotal role of ICln for cell function, which is specified below.

1.1.1.1 ICln and Regulated cell volume degrease (RVD)

The regulation of cell volume is essential for cell function. Metabolic rates and rates of substrate transports are influenced by the concentration of enzymes and intracellular ions. Other important cell events like proliferation, apoptosis, hormone and transmitter release, excitability and contraction, migration, pathogen-host interactions are also influenced by changes in cell volume.

Under hypo-tonic conditions animal cells react against swelling by releasing intracellular K^+, Cl^- and organic (e.g. inositols, taurine or L-arginine) solutes accompanied by water to keep their cell volumes constant. This homeostatic response is termed *regulatory volumes decrease* (RVD). The biophysical characteristics of the Cl^- current activated upon hypo-tonic challenge, $I_{Cl,swell}$ are remarkable constant in different cell types [130]. The current reaches a steady state level of activation after few minutes, presents voltage dependent inactivation at potentials above E_{Cl}, is inhibited by a wide variety of compounds, including conventional anion transport inhibitors, and can be blocked by extracellular nucleotides such as ATP. $I_{Cl,swell}$ exhibits the permeability $I^- > Br^- > Cl^- > F^-$ (Eisenman type I sequence). The activation of $I_{Cl,swell}$ depends on the presence of intracellular ATP, maybe because to prevent the loss of valuable intracellular organic osmolytes during starvation. No ATP hydrolysis is necessary, as known from patch clamp experiments with nonhydrolyzable analogs. High concentration of intracellular chloride shift the set point of activation to larger cell volumes. The mechanism by which cell swelling activates $I_{Cl,swell}$ is still unknown.

The Identification of the $I_{Cl,swell}$ related chloride channels and their regulatory proteins is currently the focus of extensive research because changes in the chloride permeability of cells are the basics of various human inherited diseases, as Dent's disease, osteoporosis, and Bartter syndrome [177], [191]. Activation of cell swelling dependent chloride channels in the heart leads to a shortening of the action potential, which could then lead to cardiac arrhythmia [44].

Initially, ICln was presented by Paulmichl and co-workers to be the swelling-induced chloride channel itself [137]. When expressed in *Xenopus oocytes* ICln gave rise to an outwardly rectifying current, that was blocked by addition of nucleotides to the extracellular solution and inactivated by depolarizing voltages. Although the ICln associated current did not require swelling for activation, its characteristics was found to be similar to the currents elicited in cultured cells under hypo-osmotic conditions [67]. For verification of its putative channel function, purified ICln was reconstituted in lipid bilayers and reported to generate permeability for both, anion

and cation in a Ca^{2+} (increased Ca^{2+} concentration shifts the relative permeability towards anions) depending manner. The open probability was found to be sensitive against nucleotide analogues. In contrast, Strange and co-workers [100] were unable to detect an anion channel activity of reconstituted ICln. Hydrophobicity analysis indicated that ICln lacks trans-membrane helices believed to form the pore of most vertebrate ion channels. Instead a homo-dimer model of ICln was designed and suggested by Paulmichl and co-workers [44], where each monomer contained four β-strands that contribute to the formation of an eight-stranded, anti-parallel β barrel trans-membrane pore, similar of that of bacterial an mitochondrial porins. Furthermore, a putative nucleotide binding site was identified in the second β-strand of each monomer, which, however, exhibits a poor homology with known nucleotide binding motifs.

Basing on biochemical and immunofluorenscence studies, which indicated a primarily localization in the cytoplasm under normal condition rather than next to the plasma membrane as commonly expected for an ion channel, [88] proposed that ICln more likely acts as channel regulator. The author has shown, that oocyte swelling gave rise to an endogenous $I_{Cl,swell}$ that superficially resembled the current induced by heterologous expression of ICln. Furthermore, although generally both currents closely resemble each other [187] point out some clear difference between them in respect to anion permeability, voltage depending block by extracellular cyclo AMP in case of ICln (which is independent for $I_{Cl,swell}$), or voltage depending inactivation (in contrast to ICln pH sensitive for $I_{Cl,swell}$). [19] demonstrated that the expression of an unrelated protein, $ClC-6$, induces an analog current as found by expression of ICln. Thereon, it was concluded that the expression of certain endogenous protein gave rise for an outwardly rectifying current different from endogenous $I_{Cl,swell}$.

However, specific knock down of the ICln protein using ICln-specific antibodies [118] or antisense oligodeoxy-nucleotides [54] lead to a significant reduction of $I_{Cl,swell}$, suggesting a direct link between ICln and cell volume homeostasis. As mentioned above, hypo-tonic stress causes a translocation of ICln from cytosol to the membrane fraction. As response against cell swelling the cytoskeleton alters (particularly F-actin) in a way by strengthening its network next at the cell membrane, which could explain the positional translocation of ICln by its capability to bind cytoskeletal components as actin.

That findings supports a rather regulatory role of ICln in volume homeostasis, although [55] and [43] disagree and affirmed, basing on their experiments of reconstituted ICln, that ICln could be a potential channel candidate of $I_{Cl,swell}$. They refer to the common features of ICln with $I_{Cl,swell}$ as rectification, the dependence of relative permeability between K^+/Cl^- on calcium

and particularly to the nucleoside sensitivity. Furthermore, they shown that reconstituted ICln is permeable to the osmolyte taurine.

Beside ICln, several other proteins as P-Glycoprotein, $ClC-2$, and $ClC-3$ were initially suggested to be responsible or part of $I_{Cl,swell}$ *in vivo*[129], but all these candidates failed [157]. However, identification of the chloride channel responsible for $I_{Cl,swell}$ is a challenging problem. For example, expressions cloning and characterization of heterologously expressed ion channels would need a cell system with minimal or no background expression of the channel of interest. Here, the situation is particular complicated because $I_{Cl,swell}$ is a general feature of most, if not all, cells that are used as expression system. The most common used expression cloning systems *Xenopus lavis* oocytes offers numerous background chloride channels, some of those seem to be activated by expression by almost any protein. Furthermore, although purification of the putative channel protein and subsequent reconstitution in bilayers efficiently help to avoid contamination with other membrane fractions, this entails loss of the natural environment and possible conformational changes of the fold topology. In addition, a specific, high affinity inhibitor(s) for $I_{Cl,swell}$ is not available so far.

1.1.1.2 ICln in platelet activation

Platelets, or thrombocytes, are involved in primary homeostasis, were they get activated by binding to collagen which is exposed when the endothelial blood vessel is damaged. Once activated, they release a number of different coagulation factors and platelet activating factors. The platelets adhere to each other and to the endothelial cells of the blood vessel wall via integrins by forming a homeostatic plug within seconds after an injury. The platelet specific integrin $\alpha_{2b}\beta_3$ is the most studied of all integrins and has been utilized as model for exploring integrin activation and ligand interaction. Many integrin binding ligands encompassing cytosceletal protein as talin or α-actinin, cytosolic proteins and membrane proteins such as $CD9$ and $CD153$ have been identified. These ligands bind preferentially to the activated form of integrins. As response to cellular stimulation (induced by contact with collagen, thrombin (via $PAR-1$ trombin receptor), binding to ADP via platelet purinergic $P2Y$ receptors [1] or other activating factors) the large extracellular part of integrin $\alpha_{2b}\beta_3$ conformationally changes presumably coordinated by the endogenous thiol-isomerase activity of the β-subunit [131]. However, of particular interest in the activation of $\alpha_{2b}\beta_3$ integrin is the role of its short cytoplasmic domain containing

[1] In platelets binding to ADP leads to mobilization of intracellular calcium ions via activation of phospholipase C, a change in platelet shape, and probably to platelet aggregation.

the membrane adjacent KVGFFKR and KLLvXiHDR (v = V, I, or M; i = I, F, or L; X unconserved residue) motifs of the α_{2b} and β cytoplasmatic subunits, respectively, which are highly conserved among the integrin family (e.g. K*L*GFFKR in α_2). [68] demonstrated that charge reversal mutations modulates the $\alpha_{2b}\beta_3$ activation state by disrupting a potential salt bridge between residues R995(α_{2b}) and D723(β_3) of the the membrane-proximal cytoplasmic subunit domains. Mutation of either residues to Ala results in a constitutively active $\alpha_{2b}\beta_3$ complex, whereas cross-substitution of these residues to reverse the charge return the system to rest. NMR studies [193] of two synthetic peptides comprising the membrane-proximal regions of the integrin cytoplasmatic tails (M987-W*K*V*GFFK*-R995-N-R997 of α_{2b}, and *K*716-*LL*ITI*H*-*D*723-*R*KEFAKFEEERARAKW-D740 of β_3; highlighted residues = highly conserved residues, italicized residues = residues implicated in salt bridge formation) revealed a α-helical structure of these fragments in the complex. Numerous NOE contacts were observed between highly conserved aliphatic residues of the β_3 peptide and aromatic residues in the conserved α_{2b} region, verifying the dimerisation.

Exploiting a large recombinant expression library (encompassing about 37200 human proteins, [94]) for detection of novel integrin binding partners, 13 proteins were found to interact specifically and significantly to a biotin tagged peptide containing the α-integrin cytoplasmatic binding-motif (KVGFFKR). From that set, ICln was found to exhibit the highest affinity. Furthermore, as determined by PCR analysis of platelet mRNA, ICln was examined to be expressed itself in platelets. A direct regulatory role of ICln in platelet activation was suggested by using acyclovir (2-amino-9-((2-hydroxyethoxy)methyl)-1H-purin-6(9H)-one, $C_8H_{11}N_5O_3$), a specific inhibitor of the ICln mediated chloride efflux [54], which results in a dose-dependent inhibition of platelet aggregation and integrin activation, but had no effect to integrin-independent platelet activation, such as stimulation by thromboxane.

1.1.1.3 ICln in snRNP biogenisis

The removement of introns from *pre*-mRNA is catalyzed by the splicosome, which is build up by the ordered interaction of the $U1$ and $U2$ small nuclear ribonuclear proteins (snRNPs), the $U4/U6/U5$ tri -snRNP particle, consisting of four snRNA moleculs ($U1$, $U2$, $U4$, and $U5$; abbreviated as U RNA) and at least 25 proteins, a set of (per date) seven core snRNP proteins (Sm proteins), and a set of snRNP-specific proteins with the pre-mRNA [2]. The Sm proteins are denoted as B'/B, $D1$, $D2$, $D3$, E, F, and G, and share a conserved sequence motif consisting

of two segments, $Sm1$ and $Sm2$, separated by a spacer region of variable length. The crystal structures of the two dimers [74] complexes B/$D3$ and $D1$/$D2$ exhibit a identical Sm motif for all four proteins and a virtually identical arrangement of the dimer partner proteins to each other. Furthermore, it was demonstrated that the Sm motif is sufficient and necessary for the formation of stable dimers. From the relative orientation of the Sm monomers in the dimer complex and their interaction contacts, a model was proposed, in which the Sm proteins oligmerize to form a seven-member ring [1]. This model was found to be in good agreement with the observed morphology and size of the Sm core under the electron microscope. The assembling of Sm proteins onto U RNA were shown to require the interaction with a complex consisting of SMN (survival of motor neurons), which provides the binding interface, and $SIP-1$ (SMN interacting protein 1) [105]. Binding of SMN to Sm proteins is enhanced by modification of specific arginine residues in the Sm proteins $D1$ and $D3$ to symmetrical dimethylargenines (sDMA) [116], assumed to represent a (negative) regulator step at the post-translational level. While the snRNP-specific proteins were related to splicing reaction, Sm proteins are only known to be involved in the biogenesis of snRNPs. $U1$, $U2$, $U4$, and $U5$ snRNAs are transcirpted in the nucleus by RNA polymerase II and offering a monomethyl cap. After export to the cytoplasm the RNAs interacts with the Sm proteins to form the Sm core. Subsequently, the mono-methyl cap is hypermethylated to a tri-methyl cap. The assembled Sm core and the trimethyl cap modification provide the import signals for the following translocation of the U snRNP complex to the nuclear [141].

Several Sm proteins were found to interact with ICln. [141] demonstrated that this interaction inhibits the assembling of Sm protein to U RNA and interferes with snRNP biogenesis. ICln inhibits SMN-binding to SmB/B', $SmD1$, and $SmD3$ proteins, thus impeding the formation of the Sm core. In addition, ICln was found to be a component of sDMA modification complex, the methylosome, further consisting of the Sm proteins, the methyl transferase $PRMT5$, and two so far uncharacterized factors. It was suggested that ICln guides newly synthesized Sm proteins to the methylosome, where B/B', $D1$, and $D3$ receive the sDMA modification [116].

Based on sequence homology, Sm like (LSm) proteins were identified that contain the Sm sequence motifs, $Sm1$ and $Sm2$, found in Sm proteins [112]. According homology to human Sm proteins, human LSm proteins were classified into seven subtypes, $LSm2$ to $LSm8$. They were found to exhibit a heptamere structure similar to the Sm core, but in contrast also in the absence of snRNA. Paradoxically, the Sm motifs of the Sm and the LSm proteins show no obvious difference that would explain the different stability of these complexes. It was proposed that a

long, proline rich domain of SmB, which is lacking in all seven LSm proteins is responsible for that finding.

The LSm core were demonstrated to be involved in pre-mRNA splicing via interaction with the U6 snRNA, although its exact function there is unknown. However, LSm mutants results in deficient splicing, which correlates with decreased $U6$ snRNA levels. Another function of LSm proteins was shown in cytoplasmatic RNA degradation. Yeast strains defective in LSm1 to LSm7 exhibit a decreased rate of mRNA turnover [176]. In addition, LSm proteins were shown to interact with mRNA decay factors, as the Dep1p decapping enzyme, or the exonuclease $Xrn1p$.

1.1.1.4 Motivation

In order to enable a more accurate interpretation of the experimental data published so far and to relate these biological findings to the molecular level, one of the first tasks in this work was the NMR determination of the tertiary ICln structure in solution. Initial attempts to crystallize wildtyp ICln failed (Fuerst, unpublished data). First, the structure of a construct encompassing residue Met1 to residue Glu149 of *Canis familiaris* (dog) wild type ICln sequence [137] was known from CD-spectra and primary sequence analysis to contain permanent secondary structure elements. Contrariwise, for the residual C-terminal region of ICln a random coil structure was predicted and a considerably lower thermodynamic stability in solution was observed. However, in order to detect possible preference for secondary structure for that part a construct containing the C-terminal residues Glu149 to His235 of ICln was analyzed. Dynamic characteristics for both the structured N-terminus and the unfolded C-terminal part of ICln were inferred from T_1 and T_2 relaxation parameter measurements.

For structure-function relationship analysis several constructs were provided by the group of Prof. Paulmichl. The putative interactions with miscellaneous ligands, as Ca^{2+}, AMP, different phosphoinositols, a 11 residue peptide containing the α-integrin cytoplasmatic binding-motif or with the protein $LSm4$ (basing on sequence homology LSm4 is related to $SmD3$, which was shown to interact with ICln) were studied. In addition, the structural behavior of ICln in different solution environments as with CHAPS in micellar concentration or the stability increasement of its C-terminal part in the presence of the osmolyt *L*-argenine was examined.

1.1.2 Cyclophilin D

1.1.2.1 Cyclophilins - occurrence and cellular functions

Cyclophilins (Cycs) comprise a highly conserved, ubiquitous ensemble of proteins, which form together with the FK506 binding proteins (FKBP) the family of immunophilins. Cycs range in size from 15 to 40 kDa and has already been isolated from organisms as diverse as *Neurospora* [181], *Escherichia coli* [61], *Saccharomyces cerevisiae* [82], plants, and have been identified in every tissue studied in humans [155]. Several Ca^{2+}-dependent and Ca^{2+}-independent cellular signaling cascades within immune and neuronal cells are regulated by cyclophilins.

Some of best functionally characterized cyclophilins are Cyc A (predominantly found at the cytosol), Cyc B, Cyc C and Cyc40 (endoplasmatic reticulum), Cyc D (mitochondria). Cyclophilins possess peptidyl prolyl *cis-trans* isomerase or rotarmase activity, which can be inhibited by binding of immunesuppressant ligands as cyclosporin A. Some proteins are found that requires cyclophilin assistance to complete folding, such as collagen, rodopsin and pseudoazurin. Cyclophilins have many clinical applications. In breast carcinomas [197], lung cancer [81], and certain liver cancers [120] a dramatic increase in the expressions level and the half-life of certain cyclophilin mRNAs was found when compared to non-cancerous tissues [90]. A mitogenic role for that cyclophilin was suggested [120]. Cyc A is incorporated into human immunodeficiency virus type 1 (HIV 1) via contact with the gag protein, which contains information sufficient for the assembly and release of virions. Genetic or pharmacologic disruption of Cyc A incorporation causes a quantitative reduction in virion infectivity [14]. Cyc A is also the target for immunesuppressant in the treatment of transplant patients. It binds with high affinity cylcosporin A, an 11-member cyclic peptide. The CsA-CypA complex interact with calcineurin, a Ca^{2+}-calmodulin activated protein phosphatase. One of its substrates is the phosphorylated form of the transcription factor NFAT, which activates the transcription of many T-cell specific genes only in the non-phosphorylated state, leading to immunosuppression [161]. Cyc 40 interact with the $hsp90$-complex and is involved in heat shock response. Some nonhuman cyclophilins can act as allergens [153]. CycA, CycB and CycC exhibits Ca^{2+} or Mg^{2+}-dependent nuclease activity and were suggested to act as apoptotic nucleases [121] (in a stage of apotosis, where nucleus-lamina is resolved). CycD is involved in opening of the mitochondrial permeability transition pore complex (PTPC) which is thought to be sufficient in triggering apoptosis. CsA, by binding to CypD, can prevent this process. It has been shown that non-immunesuppressant FKBP ligands stimulate neurite growth in culture and promote regeneration of peripheral and

central nerve fibers *in vivo* ([172], [202]).

1.1.2.2 Cyclophilin D and Apotosis

Mitochondrial functional complexes providing contact sites between the mitochondrial inner and outer membrane are though to be central to apoptotic processes [27]. The contact sites are formed by the inner membrane adenine nucleotide translocase (ANT) and the outer membrane voltage dependent anion channel (VDAC). Both perform established roles in the daily life of the cell. ANT mediates ADP-ATP exchange across the inner membrane, essential for the basic bioenergetic function of the organelle in supplying ATP to the cytosol. VDAC, or mitochondrial porin, is a large water-filled pore that allows low molecular weight solutes up to 1.5 kDa to permeate freely across the outer membrane and to gain access to the solute-specific transport systems of the inner membrane, which is almost impermeable to allow generation of the electrochemical gradient ($\Delta\psi_m$) used for oxidative phosphorylation. The ANT/VDAC-complex binds a range of proteins such as hexakinase, creatinekinase and enzymes of lipid transfer. It also offers a high affinity to cyclophilin D (CycD) (see figure 1.2).

Furthermore, anti-apoptotic homologous of the Bcl-2 family (Bcl-xL, Bcl-2) as well as the pro-apoptotic proteins Bax and Bak are able to interact directly with ANT, as indicated by co-purification, co-immunoprecipitation studies and yeast two hybrid screening [111]. The pro-apoptotic proteins can shuttle between a non-mitochondrial localization, while the anti-apoptotic factors reside in the mitochondria and stabilize the barrier function of the mitochondrial membranes.

VDAC, ANT and CycD are all that is required to form the mitochondrial permeability transition (PT) pore involved in triggering apoptosis [26].

Under physiological conditions the periodic reversible opening of the PT pore allows for release of Ca^{2+} from mitochondrial matrix, thereby participating in Ca^{2+} homeostasis, as well as ATP/ADP exchange between matrix and cytosol. Irreversible permeability transitions leads to mitochondrial autophagy and cell death. This event short called permeability transition involves a sudden increase in permeability of the inner mitochondrial membrane to molecules up to ≤ 5 kDa.

The probability of PT pore opening is regulated by several factors and environment variables. Pore antagonists shift the gating potential to more negative values (physiological = 200mV, with negative potential inside), favoring pore opening, whereas pore antagonists favor its closure.

The PT pore is closed at acidic mitochondrial matrix pH, whereas alkalinization leads to opening. Matrix Ca^{2+} ions increases the probability of channel opening. Matrix Mn^{2+}, Mg^{2+} and external bivalent metal ions all decrease this probability. The oxidation/reduction state of pyridine nucleotides ($NADH/NAD^{+}$, $NADPH/NADP^{+}$) also influence the pore complex. Oxidation of these species increases permeability of the pore, whereas the endogenous ligand ADP as well as glucose and creatine inhibit permeability transition [201].

Full-blown permeability transition leads to equilibration of ions within the matric and intermembrane space of mitochondria followed by a loss of the mitochondrial transmembrane potential $\Delta\psi_m$, thus interruption of ATP synthesis, uncoupling of the respiratory chain followed by hyper generation of cell toxic superoxide anions. it's also accompanied by a volume dysregulation of mitochondria due to hyperosmolality of the matrix, which causes the matrix space to expand. Because the inner membrane with its folded cristae possesses a larger surface area than the outer membrane, this matrix volume expansion cause outer membrane rupture, releasing caspase activating proteins located within the inter-membrane space to the cytosol. Caspases are a family of cysteine proteases that cleave proteins after aspartic acid residues and which are key mediators of apoptosis.

Interestingly, the opening of the permeability transition pore were shown to be inhibited by binding of cyclosporin A and derivates to cyclophilin D, thus preventing apoptosis [175]. This finding is of particular interest in the drug development for the treatment of neurodegenerative diseases, as described below.

1.1.2.3 Cyclophilin D as drug target in neurodegenerative diseases

Neurodegenerative diseases as Parkinsons disease (PD) and Alzheimers disease (AD) are often accompanied by mutations in the mitochondrial genome [192]. Defects in nuclear genes coding for respiratory chain components in fractions of both, AD and PD patients, were found. Transfection of such mutant mitochondrial DNA to cells lacking *mt*-DNA yields cytoplasmic hybrids, which are more susceptible than control hybrids (bearing normal *mt*-DNA) to the dissipation of $\Delta\psi_m$ by inhibitors of the respiratory chain or to elevations of cytosolic Ca^{2+}, a inducer of PT pore opening. Remarkable, these alterations disappear by treatment with ligands to Cyclophilin D [22]. This finding suggest that a constitutive increase of PT pore activity is occurring, presumably as a result of enhanced generation of free reactive oxygen species (ROS) by the deficient electron transport system [22].

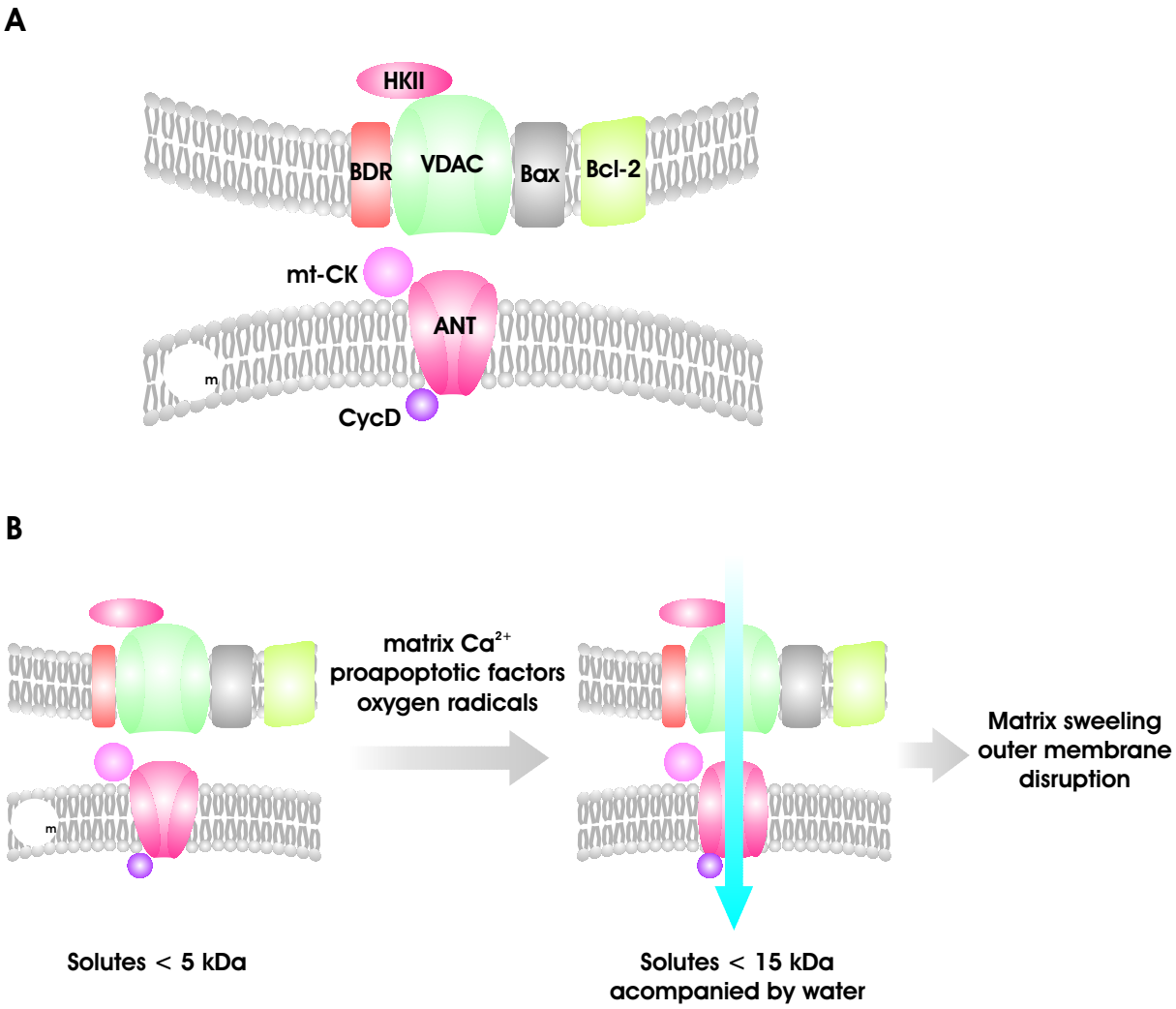


Figure 1.2: A model for the mitochondrial permeability transition pore complex. (A) The permeability transition pore complex (PTPC) provides a contact site between the inner and the outer mitochondrial membrane and involves several transmembrane proteins: the adenine nucleotide translocator (ANT), the voltage dependent anion channel (VDAC) and the peripheral benzoadizepine receptor (BDR). it's also associated with other proteins such as hexokinase II (HKII), mitochondrial creatine kinase (mt-CK), the peptidyl-prolyl isomerase cyclophillin D (CycD). The ANT is responsible for exchange of ATP and ADP on the inner membrane. The VDAC functions as a nonspecifi c pore allowing diffusion of solutes up to 5 kDa. The exact function of the PDR is unknown. (B) VDAC, ANT and CycD are all that is required to form the mitochondrial permeability transition pore. Cellular signals triggering apotosis arouse a sudden increase in permeability of the inner mitochondrial membrane to solutes up to $\leq$ 15 kDa. Ensuing enter of water and solutes into the mitochondrial matrix cause matrix swelling and outer membrane disruption. (IMM inner mitochondrial membrane, OMM outer mitochondrial membrane, $\Delta\psi_m$ electrochemical proton gradient).

Enhanced ROS generation can also pandered by mutations in nuclear genes encoding proteins, which are imported to mitochondria. A decreased expressions level of the mitochondrial protein frataxin, which mediates the efflux of excessive Fe^{2+} from the matrix as well as knock out of mitochondrial manganese-dependent superoxide dismutase is followed by increasing of ROS [144].

Several reasons have been suggested why preferentially neuronal cells are effected by perturbations of mitochondrial function.

Neuronal cells are more susceptible than other cells to inductions of PT pore opening [204]. Certain regions of the brain (substantia nigra) exhibit an increased level of iron, which may lead to more sensitivity against oxidative stress. Neuronal cells may be particularly irritable to perturbations of bioenergetic function, with a comparatively high threshold level below which perturbations of ATP production will cause cell death [186].

However, as mentioned above it has been shown that suppression of the mitochondrial permeability transition results in neuroprotection. Small molecular weight inhibitors of Cyclophilin D allowing brain penetration are thus highly desirable, since they have the potential of being useful therapeutic agents for treating neurodegenerative diseases.

1.1.3 LIM1 domain from quail cysteine- and glycine-rich protein CRP2

1.1.3.1 LIM domains - classification and function

Initially, the LIM domain was derived as a histidine-cysteine double zinc finger motif within LIM homeodomain proteins [174]. To date, the family of LIM proteins rapidly increased encompassing members that lack homeodomains either by being composed almost entirely by LIM domains (LMO proteins) or containing additional functional domains of various types (e.g. kinases in LIMK)[203]. An early classification scheme was based on sequence relationship between lim domains and on the overall structure of the proteins [28]. More roughly, LIM proteins can be subdivided into predominantly nuclear LHX (LHX$_i$, i=1-8; ISL$_j$, j=1-3 ; MEC-3; BK87; HRLIM; LIN-11) and LIM-only proteins (LIM$_i$, i=1-4), and proteins with LIM motifs that display a dual cellular localization (LIM-kinase 1, LIM-kinase2, cysteine rich proteins CRP$_i$, i=1-3; lipoma prefered partner (LPP); Zyxin; Paxillin; Enigma; TRIP6; Four and a half LIM domain (FHL) proteins). Recently, a subclassification approach for individual family members was suggested that employ a hydrophopicity/isoelectric point versus residue exposure or compactness profile (calculated for each sequence) for further discrimination (Konrat, unpublished). This method

was also indicated to be useful for localization prediction of unknown LIM domains.
LIM proteins are predominantly found in humans and animals, although some has also identified in yeast and plants.
A functional hallmark of all LIM domains is presumably the mediation of protein-protein interaction by acting as multiple binding and adapter modules [203]. LHX proteins are transcription factors that bind DNA through their homeodomain and were shown to be important developmental regulators. Many of the cytoplasmatic LIM proteins , such as zyxin, CRP and paxilin, are associated with the cytoskeleton. Proteins offering only short regions beyond their LIM domains, such as CRP's and LMO's, appear to function solely as intermediate connectors to bring other complex to a complex. The larger zyxin is capable to bind via its (first) LIM domain to CRP's and to α-actinin through its proline rich N-terminal domain.
Beside forming dimers with other LIM domain, it become apparent that LIM domains can bind various structurally different protein motifs [73]. Several of that corresponding proteins (e.g. paxilin,zyxin) are components of the adhesion plaques through which cells adhere to the extracellular matrix. The LIM domain protein ENIGMA was shown to bind the receptor tyrosine kinase RET and some isoforms of protein kinase C (PKC) in a LIM-domain dependent manner.

1.1.3.2 Structure of LIM domains and motivation

The solution structure of several LIM proteins have been determined by NMR spectroscopy including CRIP, the N- and C-terminal LIM domains of chicken CRP1 (commonly referred as LIM1 and LIM2) and of quail CRP2.
The LIM motif is composed of two cysteine-rich zinc finger structures separated by a two amino acid spacer and conforms to the general consensus sequence $CX_2CX_{16-23}HX_2CX_2CX_2CX_{16-21}CX_2$ [174]. The two zinc finger represent separate structural entities that held together in an elongated configuration via a hydrophobic interface. The exact orientation of both subdomains to each other as well as their packing tightness fairly depends on the length, distribution and density of the hydrophobic sidechains. Each LIM domain contains two tetrahedral Zn^{2+}-coordinating sites of the CCHC and CCCC type, respectively ([119], [87]). Both LIM domains in CRP1 and CRP2 were found as structurally independent and connected by an apparently unstructured linker region of $\sim$ 50 amino acids. Furthermore, the structural and dynamical properties of the full length proteins were essentially identical to the isolated LIM domains.
Despite a close structural similarity between the C-terminal CCCC zinc finger segments and the

GATA-1 transcription factor, until now there is no distinct evidence for DNA-binding capabilities of LIM domain proteins.

So far, not any solution structure of a LIM protein complex is available.

One aim in this work was the refinement of the amino terminal LIM domain (LIM1) from quail CRP2, which was initially determined using a ^{15}N-labeled sample only [84]. By employing a full set of standard triple resonance experiments and simultan ^{13}C-^{15}N 3D NOESY a rarefied structure was generated that was subsequent used by Paul Schanda and Elisabeth Dowler (Prof. Konrat group, unpublished results) for rotational diffusion tensor determination. This method analyzes the local diffusion coefficients derived from relaxation rates in order to calculate the diffusion tensor assuming isotropic, axially symmetric, and asymmetric diffusion [33].

1.1.4 Brain abundant signal protein 1 - BASP1

The BASP1 (Brain abundant signal protein 1, other historical names are $CAP-23$ and neuronal axonal membrane protein $NAP-22$) is an acid soluble (5% $HClO_4$) protein found to be abundant in nerve endings. It's thought to act as signal protein participating in neurite outgrowth and plasticity [123]. BASP1 can be phosphorylated *in vitro* by protein kinase C (PKC) at residue serine 6. This modification is inhibited per binding of calmodulin in a calciun depending manner. Via N-myristoylation BASP1 is able to anchor to plasma membrane. Elong its primary sequence further hydrophobic segments are virtually absent. BASP1 contains at least two PESTD sequences, shown to be characteristic for short living proteins [151]. A primary sequence comparison of a set consisting of 12 proteins offering various different cellular functions and a intracellular half-lives less than 2 hours has revealed commun high local concentrations of amino acids proline (P), glutamic acid (E), serine (S), threonine (T) and to a lesser extent aspartic acid (D), often flanked by a positive charged residue (in BASP1 that is residue type lysine). Conversely, deleting of that PESTD regions remarkable correlates with an increased intracellular stability.

From these findings it was concluded that PESTD motifs reduce the half-lives of proteins dramatically and hence, that they target proteins for proteolytic degradation. Conservation of proteins with PESTD sequences among different species supports their functional significance (see figure 1.3).

BASP1 was also found in some non nervous tissue, as spleen, thymus, kidney and testis, where its function is poorly understood. However, BASP1 was identified as transcriptional co-

Figure 1.3: Multiple sequence alignment of BASP1: Gallus gallus (chicken), Bos taurus (bovine), Homo sapiens (human), Mus musculus (mouse), and Rattus norvegicus (rat). Where necessary, gaps (shown by points) were inserted to indicate the homology. Software applied for alignment and representation as well as the color code is analog to fi gure 1.1. The three conserved PESTD segments are indicated by gray bars. The sequential hydropathic character (as given by the hyd line under the alignment) was calculated according [93] using a window size of 3 (pink=hydrophobic, gray=intermediate, green=hydrophilic).

suppressor for the Wilm's Tumor suppressor zinc finger protein $WT1$ [20]. WT1 was shown to regulate genes for growth factors and regulators of cell division. Wilm's tumor is the most common malignancy of kidney at childhood age.
In all that tissue, BASP1 was found to occur along with a small set of its N-terminal fragments devoid of C-terminal parts of different length. Similarly to full length BASP1, all these fragments were found to be myristoylated. Beside human, this finding was also verified for other species as rat, bovine, and chicken, and a physiological significance was suggested [200].

1.1.5 Osteopontin

1.1.5.1 Molecular origin, structure and functionality

Osteoponitin (OPN) is a hydrophilic phosphoprotein found in numerous tissues (gland, kidney, brain, bone, smooth muscle, and immune organs) expressed by several immune cells (T-cells, B-cells, natural killer cells, platelets, macrophages), vascular smooth muscle cells, and diverse epithelial cells. OPN is secreted in all body fluids (bile, urine, milk, and sweat), and function to promote cell adhesion, to facilitate cell migration or cell survival via interaction with integrins and $CD44$ variants (see below). It's a member of the small SIBLING (Small Integrin Binding Ligand N-linked Glycoprotein) family of genetically related proteins [40]. The integrin attachment motif GRGDS is completely conserved among species (see figure 1.4). A high degree of sequence conservation is also observed for the neighboring protease hypersensitive thrombin cleavage site (located seven residues C-terminal to the RGD sequence, red bar in figure 1.4).
OPN is postranslationally modified by O-glycosylation, sulfation, and serine/threonine phosphorylation (most of them located at the recognition motif for mammary gland casein kinase (Ser/Thr-X-Glu/Ser(P)/Asp)) [80]. Phosphorylation occurs to a variable extent up to 28 sites depending on the cellular origin. It can also be covalently incorporated into extracellular matrix by transglutaminase cross-linking to lysine residues of fibronectrin. OPN binds strongly to hydroxy apatite [47]. The mineral binding poly-aspartarte (pAsp) region only differs in the number of consecutive aspartic acid residues. Via that pAsp region OPN serves as an attachment protein linking cells to the bone mineral. Thereby, for example, OPN offers a regulative function in crystal growth and nucleation by recruitment and migration of osteoclasts, whose efficiency strongly depends on the phosphorylation level of OPN. OPN is also a substrate for the extracellular matrix metalloproteinases MMP-3 and MMP-7 [150]. As with thrombin treatment,the resulting fragments of OPN exhibit an enhanced ability to stimulate cell adhesion and

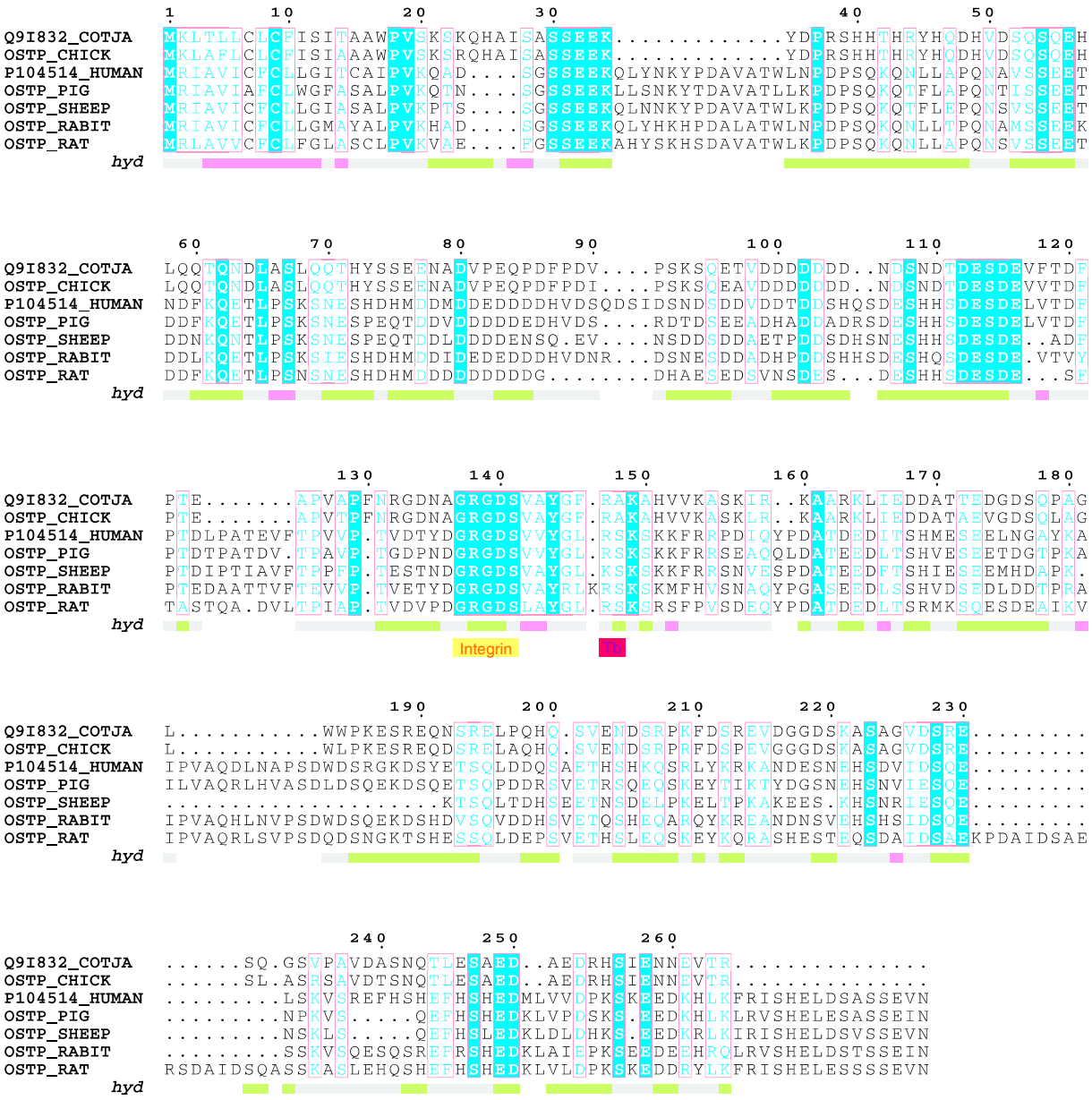


Figure 1.4: Multiple alignment of osteopontin sequences:Coturnix coturnix japonica (japanese quail), Gallus gallus (chicken), Homo sapiens (human), Sus scrofa (pig), Ovis aries (sheep), Oryctolagus cuniculus (rabbit), Rattus norvegicus (rat). The integrin attachment motif GRGDS and the thrombin cleavage site are indicated by yellow and red bars, respectively. Software applied for alignment and representation as well as the color code is analog to fi gure 1.1.

migration.

Binding to members of $\alpha_v\beta$ integrin and CD44 receptor families allows OPN to mediate adhesive cell matrix interactions and activates cellular second messengers in signal transduction pathways [190]. Integrins are heterodimeric cell surface glycoproteins encompassing non-covalently associated α and β subunits. Both subunits are integral membrane proteins attached to the plasma membrane through a single transmembrans α-helix. Integrin ligand binding controls cell growth , differentiation, gene expression, apotosis and motility. OPN and the widely expressed integrin receptor $\alpha_v\beta_3$ were shown to enable cellular adhesion of B cells and platelets, resorption and migration of osteoclasts, as well as migration and adhesion of smooth muscle cells [30]. This binding is enhanced by Mg^{2+} and Mn^{2+}, and inhibited by Ca^{2+}.

The cell-surface CD44 glycoproteins are involved in cell-cell interactions, cell adhesion and migration [13]. Beside chondroitin sulfate and hyaluronic acid CD44 binds to a set of extracellular matrix proteins (as collagen, fibronectin and lamin) and matrix metalloproteinases (MMPs), and exists in multiple isoforms of the extracellular domain (denoted as CD44vx, x = 6 - 15). Standard CD44 (CD44s) is involved in maintenance of the spatial structure of tissue and is not able to interact with OPN, whereas its variant isoforms bind independently each of the two OPN fragments generated by thrombin. The interaction of OPN with CD44 has been associated with the migration of macrophages and tumor cells (see below).

OPN exhibits a high number of negatively charged residues (predominantly aspartic amino acids) and a lacking of extended hydrophobic segments throughout the primary sequence. Initial examination of NMR ^{1}H proton spectra [40] as well as study using attenuated total reflection infrared spectroscopy [47] suggests an intrinsic random coil conformation of OPN. This is also indicated using secondary structure prediction methods.

1.1.5.2 Role in pathological situations

OPN from kidney epithelial cells was identified as a component of kidney stones and its expression is up-regulated in stone formers [115]. While free OPN inhibits in vitro the growth of calcium oxalate and the aggregation of hydroxy apatite crystals (which are also present at a significant fraction in renal stones) by its affinity for Ca^{2+}, immobilized OPN was shown to stimulate crystal adhesion and aggregation (where OPN-Ca^{2+} complex serves as seed crystal), whereas phosphorylated form of OPN was shown be most effective. OPN was also found as part of atherosclerotic plaques and may be a mediator of arterial neointima formation as well

as dystrophic calcification that often accompanies this process [170]. OPN is involved in the inflammatory and immune response. Secreted by activated T cells, it induces macrophages to migrate to sites of inflammation [31]. Both, the extent of calcification inhibition and the ability to chemoattract macropages strongly correlate with the magnitude of OPN phosphorylation. OPN deficient mice were shown to exhibit a strongly reduced T-cell mediated immunity to viral and bacterial infection [4], a disability which is partially restored upon purified OPN injection. Transgenic mice studies also suggest a involvement in the proliferation of B1 lymphocytes and in enhanced production of immunglobulines [69]. Because of increased OPN expression levels in a wide spectra of malignancies (lung, breast, gastric, prostate, ovarian, and uterine cervical tumors), OPN act as marker for cancer diagnosis. OPN was directly related with a elevated metastatic potential. Cells from OPN knock-out mice exhibit an impaired colony formation in vitro and a slower tumor growth in vivo in comparison with tumors in wild type mice [3]. Impeding of OPN expression using anti-sense RNA resulted insuppression of both in vitro and in vivo murine colon adenocarcinoma metastatsis [8].

Interestingly, tumor derived OPN from various cell types analyzed per date appear to be less phosphorylated compared to the host produced OPN form [190]. Hence, less phosphorylated OPN is assumed to be more effective in cancer progression either by passive protecting of the tumor cells from immune response and apoptosis or by a decreased specificity of protein anchorage.

One key event in tumorigenesis and metastasis is the degradation of the basement membrane and interstitial matrix, which is mediated by proteolytic activity of matrix metalloproteases (MMP) and the plasminogen-activator-plasmin system (see figure 1.5). Binding of OPN to $\alpha_v\beta_3$ integrin induces phosphoinositide(PI)3-kinase/Akt-dependent NFκB activation and urokinase plasminogen activator (uPA) secretion [146]. The level of uPA expression correlates with metastatic potential in several cancers. OPN effectuate Nuclear factor Inducing Kinase (NIK)-dependent NFκB activation, which stimulate $MMP-9$ and $MMP-2$ activation, although Both $MMP-2$ and $MMP-9$ efficiently degrade native type IV and type V collagen, fibronectin, and elastin. The OPN/CD44vx interaction was shown to contribute in cell survival and motility through activation of phospholipase C γ/ protein kinase C / phosphoinositol 3 kinase (PI3) signal transduction pathway. One of its downstream targets is the serine/threonine kinases Akt, which regulates cell cycle progression and growth factor mediated cell survival. The gene copy numbers of the genes encoding for PI3 kinase and Akt in breast primary tumors were found to be higher than in normal breast epithelial cells suggesting a important role of that proteins in

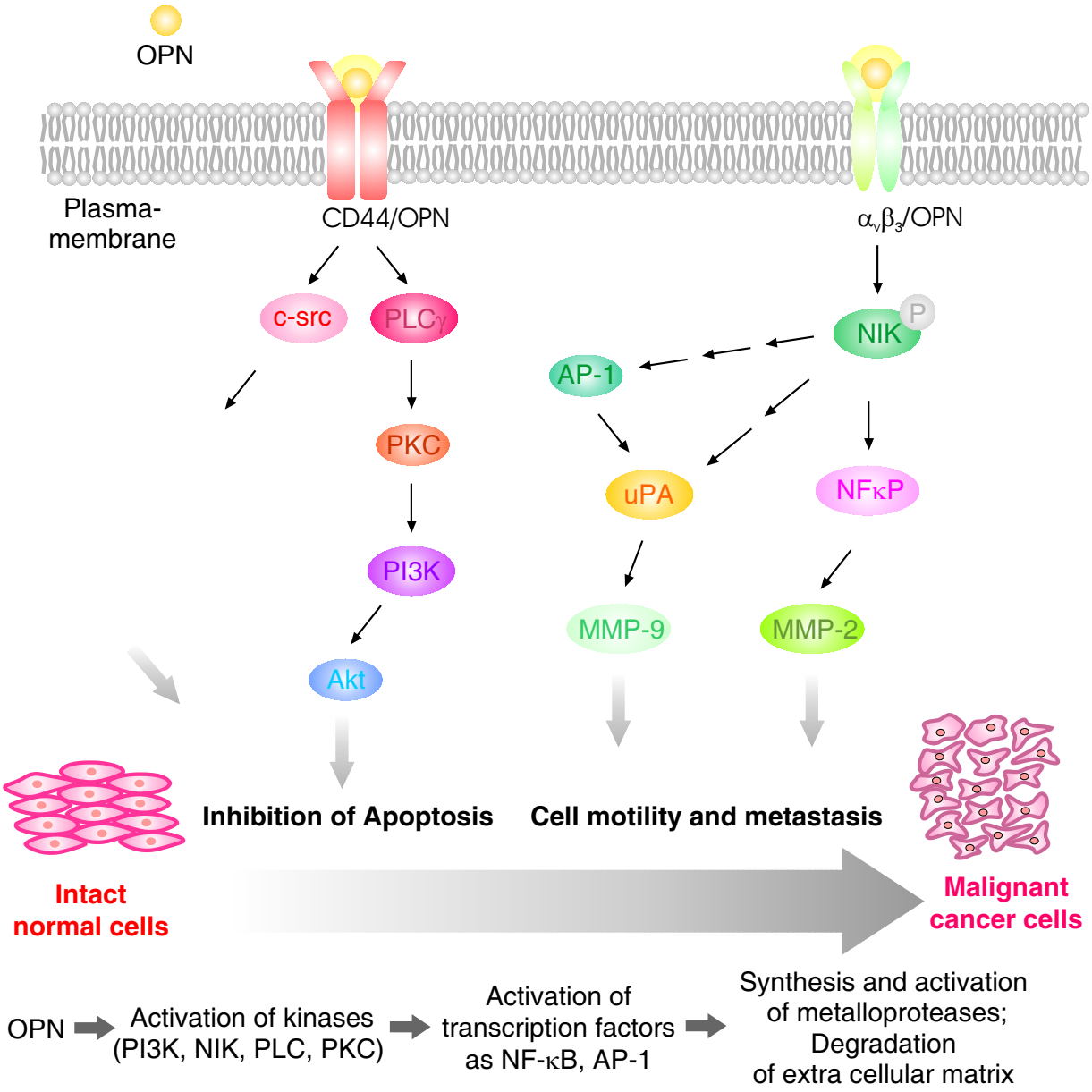


Figure 1.5: Osteopontin in tumor progression. The potential role of OPN in regulating the transformation of normal cells into malignant cancer cells is shown. Ligation of Osteopontin with either $\alpha_v\beta_3$-integrin or $CD44$ receptor (top) give rise in various cellular signaling events, resulting in survival of transformed cells and metastasis (bottom). The fi gure was generated using Corel Draw 11.0.

tumor progression [147].
In this work structural and dynamic information of that important cancer target protein in solution was derived from backbone chemical shift and relaxation data by using a construct of recombinant *quail* osteopontin containing residues His46 to Arg264 of the wild type sequence.

1.2 Towards automated fold generation

Up to now (May 2008), complete DNA sequencing and genetic mapping was performed for about 1879 viruses, 577 bacterial species and roughly 23 eukaryote organisms. The most prominent example was the successful effort to decode the human genome. By dint of experimental methods developed to accomplish such projects, sequencing and elucidation of small genomes now become routing work. However, the knowledge of the three dimensional structures of the encoded proteins is essential to determine and understand the function and course of biological processes on the molecular level. Fold homology analysis provide a helpful tool for evolutionary linking of proteins sharing marginally homologous primary sequence and different function [66]. As a result of this, structural genomics projects were initialized, which aim to determinate the 3D structure of all proteins by x-ray crystallography, NMR spectroscopy or computational approaches such as homology modeling. .
However, in contrast to a high number of deposited protein sequences (356,194 entries in uniprot/swissprot sequence database), only about 49,300 entries of protein structures in the protein database (PDB) are existent so far, most of them related to highly homologous or even identical proteins. Apparently from deposition statistics, the large cap between already solved protein structures and the number of known protein sequences will continue to increase. Thus, a massive structure determination effort is required to achieve the ambition of structural genomics. Although x-ray crystallography presents the advantage of virtually no molecular size limits, NMR spectroscopy provides the availability of structural data analyzed under nearly physiological conditions. Furthermore, dynamic properties of biomolecules or protein domains ($\leq$ 25 kD) of interest can be studied in detail routinely. Typically, a trained spectroscopist is occupied for several weeks to months in solving a new NMR structure. In order to speed up many of the time consuming step of the manually performed NMR structure analysis, several automated or at least semi-automated approaches have been proposed. Today, some programs thereof are already of widespread use.
The conventional procedure of NMR structure determination follows the successive steps of

sample generation, recording NMR experiments, processing, peak picking, sequential backbone and sidechain chemical shift allocation, NOE assignment and collection of other conformational restraints [196]. A suitable target have to be selected offering good solubility ($>$ 1mM), sufficient stability (at least 1-2 weeks), and negligible unspecific aggregation. NMR-techniques for structure determination need isotope enriched sample. Typically ^{13}C/^{15}N uniform labeled sample is used for data collection, although methods for spectra simplification employing specific labeling techniques become more and more common. Of particular interest are labeling methods using cell free expression systems, which allow incorporating of all possible natural amino acids offering a defined optimal labeling patterns practical for NMR [70]. Further advantages of cell free systems include higher success rate in expression of recombinant proteins toxic to normal bacterial host cell, its open nature allowing chaperons, detergents, ligands etc. to be added in order to increase the yield of folded protein, but also imparting a large potential for automated sample production. However, cell free systems are especially a cost-intensive task.

Once a set of spectra for sequence specific ^{1}H, ^{13}C, and ^{15}-N chemical shift assignment and NOESY data are acquired, spin systems can be created, which relate *intra*- and *inter* chemical shifts to distinct residues. A minimum combination of such NMR experiments could contain: 3D HNCA/HN(CO)CA, 3D HNCO/HN(CA)CO, 3D CBCA(CO)NH/HNCACB for ^{1}H, ^{13}C, and ^{15}-N backbone chemical shifts list generation. Valuably sequential NOEs could be derived from heteronuclear-edited NOESY spectra.

Then a spinsystem for residue i would comprise the following chemical shift information: $\delta^{1}\mathrm{H}_i^N$, $\delta^{15}\mathrm{N}_i$, $\delta^{13}\mathrm{C}_i^{\alpha}$, $\delta^{13}\mathrm{C}_{i-1}^{\alpha}$, $\delta^{13}\mathrm{C}_i^{\beta}$, $\delta^{13}\mathrm{C}_{i-1}^{\beta}$, $\delta^{13}\mathrm{C}_i$, and $\delta^{13}\mathrm{C}_{i-1}$, and a set $^{1}\mathrm{H}_i^N$-$^{1}\mathrm{H}_i^N$ of NOE connectivities.

Subsequently the best arrangement of all possible spinsystems to each other have to be evaluated. Practical hindrances in creating a usable spinsystem list can arise from missing peaks (incomplete picking or signal inexistent because of fast relaxation), peak pick errors in crowded spectral regions, artifacts of spectral processing, incomplete water suppression, sample impurities, noise, splitting of peaks caused by strong coupling, or the presence of multiple protein conformations.

Depending on the quality of the individual spin systems and the overall chemical shift dispersion, the fairly time consuming step of backbone chemical shift assignment was shown to can be successfully automated even in the presence of incorrect spinsystems by methods basing on a Monte Carlo like algorithm (e.g. MONTE [63], MARS [72], or [65]). These methods work

especially efficient when at least a pair set of $\delta^{13}C^{\alpha}$, $\delta^{13}C^{\beta}$ or $\delta^{13}C^{\alpha}$, $\delta^{13}C'$ are available as input.

^{1}H, ^{13}C sidechain chemical shift assignments can be inferred from HCCH-TOCSY and HCCC-TOCSY experiments in combination with ^{13}C NOESY spectra. Some assignment programs are also concerned with this more demanding problem [108], [63].

The most time consuming step in structure determination is the assignment of the NOESY cross peaks, which requires as a prerequisite the knowledge of the sequentially assigned chemical shifts of the spins from which the nuclear Overhauser effects (NOEs) are arising. NOEs are the most important but not the only experimental data which determine NMR structures. A series of different other data can be used in addition to NOEs, such as one and two bond coupling constants and chemical shifts [21]. Particular helpful are residual dipolar coupling (RDC) resulting in orientational restraints for structure calculation ([178], [132]). However, at least a minimal number of long range key NOEs have to be defined to obtain (at least) a reliable initial structure model. A serious problem in NOE assignment stems from the degeneracy of the resonances or their imitated chemical shift value accuracy in the spectra, respectively, which leads to ambiguous NOE identities. Large distortions in the resulting structures can arise from only a few NOE miss-assignments (depending on their sequential distance).

Especially for larger molecular systems, where chemical shift assignment and management of NOE ambiguities become laborious, an automated performance is desirable. An automated approach has to overcome several obstacles when starting without any prior knowledge of the structure. The number of initially assignable NOEs is typically much too low to define the fold of the protein, although at least a few valuable NOE contacts related to proton spins far apart in primary sequence can usually defined as unambiguous. Furthermore, an automated program must be able to cope with erroneously picked or inaccurately positioned peaks as well as with the incompleteness of chemical shift assignment of typical experimental data sets.

1.2.1 Algorithms for automated NOESY assignment demanding prior chemical shift assignment

Semi-automated approaches as the *ASNO* method [56] and the *SANE* method [34] were presented, which first determine a set of all possible chemical shift-based NOE assignments in compatibility with initial structure model. This structure could be a bundle of conformers calculated with a preliminary set of unambiguous distance restraints, but also a crystal structure

from a homologous protein. However, in such application one has to offer some caution against possible bias by the imported reference model. For both programs, the input data contain a list of the proton chemical shifts, a peak list comprising the shift coordinates of the NOESY cross peaks and the initial model. The user specifies the maximally allowed chemical shift differences between corresponding cross peak coordinates and proton chemical shift values used for NOE assignments, and the maximal distance d_{max} of the NOE related protons in the reference model to give rise for an acceptable NOE assignment. At that point, the *SANE* protocol provide additionally active assignment filters. First, any partial assignments already provided by the user is retained by the program, if it is part in the set of all possible assignments identified until now, otherwise a warning is reported. Secondly, a secondary structure filter is applied, which removes any NOE in user defined regions of secondary structure (known from secondary shifts and characteristic NOE patterns) non-non-compatible with distances typically found in α and β secondary elements (e.g. NOE assignments more than five residues apart in primary sequence). Third, relative NOE contributions calculated for each possible NOE assignment from the reference structure bundle are taken into account. This is originally an implementation of [128] (see below).

At the end, in both methods the user have to decide interactively about the assignment and/or temporary removal of entries in the NOESY cross peak list. With the remaining newly derived NOEs a new structure bundle is generated. The procedure of NOE assignment and structure calculation is repeated until a high quality structure is obtained. Because of the additional applied features for NOE assignment filtering, *SANE* trend to converge faster than the *ASNO* method.

Full automated iterative NOESY assignment protocols were presented by the *NOAH* method [125] and subsequent, introducing some completely new concepts, by the today widely used *ARIA* routine [128], [102]. ARIA has been interfaced initially with the structure calculations program XPLOR [164] and later with the program CNS [17], NOAH was inserted as subroutine in the programs DIANA [57] and DYANA[58].

If no initial starting model is available, both methods start with a model of randomized torsion angles. Then, all possible chemical shift based NOE assignments are listed and the corresponding minimum distances d_{min} or average distances d_{avg} between the NOE related protons from a structure bundle of lowest energy are calculated. In *NOAH*, individual unambiguous distance restraints are defined for each assignment possibility of a NOESY cross peak, thus accepting a distinct number of erroneous ones. No NOE distance restraints are derived from

cross peaks with too many assignment possibilities (typically more than two). In order to reduce the impact of this incorrect restraints an error tolerant target function ([128], see also ?) is used. After each round of structure calculation, a violation analysis is performed per determination of the fraction of structures in which a particular restraint is violated by more than a user defined threshold. If this fraction exceeds a cutoff number of model structures, the restraint is removed from the NOE list. By doing so, the routine dry to find a self-consistent network of correct assignments. In contrast, incorrect NOE restraints are expected to be randomly distributed in space, generally contradicting each other. In that way, *NOAH* intentional enters incorrect restraints, and hope to identify these afterwards. However, this concept need a rather high density of NOEs to work, in region of a low number of NOE as found at chain ends or surface loops incorrect restraints will most likely remain.

In the *ARIA* method ambiguous NOE assignments are inserted for structure determination in form of ambiguous distance restraints (ADR). Beside unambiguous derived distance restraints, ADR are used for each round of structure calculation. Following distance dependency and employing the isolated spin pair approximation (), an ambiguous NOE (assuming identical proportionality factor for all contributing protons) is defined/represented as d^{-6} summed distance $\bar{D}$:

$$\bar{D} = (\sum_{n=1}^{N} \hat{d}_n^{-6})^{-1/6} \tag{1.1}$$

This distance is restraint in analogy to unambiguous distance restraints using a soft square potential to take instantaneous occurrence of large violations into account. In structure calculation, ADCs will not lead to fold distortions as long as the correct assignment is also present among the initial assignments. However, because of the imitated information content of ADCs, incorrect assignments therein have to be eliminated in order to achieve convergence to the correct structure. In ARIA not any possible initial NOE assignment is accepted and converted into distance restraints, rather a filter criterion is applied basing on an estimate of the relative peak contributions of different assignment possibilities to a NOESY cross peak. A contribution values is calculated as:

$$C_i = \frac{\hat{d}_i^{-6}}{\sum_{n=1}^{N} \hat{d}_n^{-6}} \tag{1.2}$$

where $\hat{d}$ can be either the minimum distance d_{min} or the average distance d_{avg} extracted from a set of low energy structures. The individual contribution values are then ordered according

size, and a number N_p of largest contribution are chosen such that:

$$\sum_{n=1}^{N_p} C_n > p \tag{1.3}$$

This cutoff parameter p is varied in different iteration rounds, in general starting with values close to 1 for the first round to 0.8 in the last iteration. As smaller the value of p is chosen, the fewer cross peaks remain ambiguous. For example, if the shorter of the lowest two distances is 2.5Å, a value for p of 0.999 would exclude a second distance of 7.9Å, of value of p=0.95 a distance of 4.1Å and a value of p=0.8, a distance of 3.3Å. The remaining distances are converted as one ADC. Performing successive rounds of structure calculation, assignment filtering, and restraint generation, the iterative procedure results in a decreasing of the NOE data ambiguities, provided that the calculated structures are in reasonable agreement with the NOE assignments. Both, the *ARIA* method and the *NOAH* method work particularly efficient for improving and completing the NOESY assignment once a good preliminary structure model is available.

The *CANDID* method presented by [62] uses more elaborated criteria for the ranking and selecting of putative NOE assignments in ambiguous NOESY cross peaks. For the calculation of the relative contribution C_i of an initial assignment i to a NOESY cross peak, several weighting factors related to chemical shift matching CS_i, the presence of a symmetrical cross peak O_i (a peak at transposed position in the 3D or 4D heteronuclear NOESY spectrum with respect to the shift coordinates of the NOESY cross peak), the compatibility with the covalent polypeptide structure O_i, the convergence of 'network anchoring' subroutine N_i, and the compatibility with a structure model $\bar{D}_i$ are calculated. $\bar{D}_i$ is given as by equation 1.4. The covalent geometry (fixed bond length, bond angles, and chiralities) of the polypeptide chain give raise to upper limits for NOE observation on certain intraresidual and sequential distances. For evaluation of O_i, *CANDID* determine these conformation independent upper limits for each proton pair separated by one or two torsion angles. A key feature of *CANDID* is 'network anchoring' of the initial chemical shift based NOE assignment. This concept is, again (see *NOAH* method), basing on the fact, that correctly assigned restraints form a self-consistent network. The program determines the weighing factor N_i by searching for all atoms c connected simultaneously to both NOE i related atoms a and b, either by initial assignments of other crosspeaks or because the covalent polypeptide structure implies that the distance must be short enough to give raise to a NOE between a-c and b-c, respectively.

Then, the relative contribution C_i of an initial assignment i to a NOESY cross peak is calculated as:

$$C_i = \frac{CS_i min(T_i O_i N_i \bar{D}_i}{\sum_{n=1}^{N} CS_n min(T_n O_n N_n) \bar{D}_n} \tag{1.4}$$

The list of initial assignment are ranked according their C_i values, and any assignment is used for subsequent structure calculation, for which $C_i \geq C_{min}$. At the begin, the structure model is just a polypeptide chain with randomized torsion angles. In order to avoid the impact of erroneous picked noise peaks or peaks from spectral artifacts, which are not easy to discriminate from protein signals before a appropriable preliminary structure for 3D based filtering exist, *CANDID* perform in the first two cycles a ambiguous restraint combination of long range peaks (cross peaks whose assignments are exclusively related to proton pairs separated by at least 5 residues). Again, as long as the correct assignment is present in a ADR, (almost) no structural distortion will occur. Only long range peaks are combined because of their pronounced effect on the global fold compared to medium and short range restraints. If A, B, C, and D denotes the assignment sets for four long range peaks, then a 2 $\rightarrow$ 1 combination is defined in a way that the two assignment sets A and B are merged into one single ambiguous distance restraint A $\cup$ B, thus halving the number of long range restraints. Similarly, a 4 $\rightarrow$ 4 combination is performed by replacing the four restraints with assignment A, B, C, and D by four combined ambiguous restraint sets A $\cup$ B, A $\cup$ C, A $\cup$ D, and B $\cup$ C. This combination preserves more structural information.

1.2.2 Structure calculation without a priory knowledge of chemical shifts

A limiting factor of all automated NOE assignment methods mentioned above is that they rely on the availability an essentially complete list chemical shift list from preceding sequence-specific resonance assignment. This concerns especially the resonance assignment of the sidechains, which are involved in the most valuable long range NOEs. In principle, demanding an accurate and very time consuming inspection of the heteronuclear TOCSY and NOESY spectra, respectively. At present, automated assignment methods dealing with this topic, need largely interactive interventions of a trained spectroscopic. However, chemical shift assignment is just needed for positioning of the observed NOEs along the covalent polypeptide chain. Starting with the early 90ies, some methods have been presented that circumvent the tedious chemical shift assignment step. The underlying concept of these approaches is the fact, that NOESY spectra of a protein typically provide a large amount of distance information even in the absence

of any chemical shift assignment. As there is no association with the covalent structure at this point, the protein under investigation is treated as a gas of unconnected protons. Each of them are still characterized and labeled by their unassigned chemical shifts and their NOEs. Using that input a spatial proton distribution can be calculated. Provided that the resulting proton densities are sufficient clear, a covalent model can then be build into that atomic distributions in an analogical manner to X-ray crystallography, where the structural model is constructed into an electron density map.

1.2.2.1 The ANSRS approach

One of the first sophisticated attempts of introducing a direct method came with Per Kraulis real space ab initio ANSRS (Assignment of NOESY Spectra in Real Space) approach [89]. The program input consists of a complete but unassigned list of chemical shifts of all detectable ^{1}H^N-^{15}N and ^{1}H-^{13}C moieties and a NOESY cross peak list including knowledge of the ^{15}N and ^{13}C chemical shifts of the heavy atom directly bound to the NOE related protons (i.e. a 4D NOESY peak list). The algorithm consists of three distinct phases. First, a 3D distribution of unassigned protons is calculated using a simulated annealing protocol whose target function contain only an attractive term E_{NOE} and a repulsive van der Waal term E_{vdW}. In the second phase a list of all possible assignments for each ^{1}H^N-^{15}N or ^{1}H-^{13}C chemical shift entries is tabulated by employing a empirical database of chemical shifts. Using the inter-proton distances $\leq$ 5Å derived from the calculated structure as spatial constraints, residue type ^{1}H proton combinations are generated using a recursive combinatorial search. The spin combinations are build from the outer tip of the amino acid residue sidechain, rather than from the backbone going outwards, because the sidechain ^{1}H-^{13}C chemical shifts tend to be more specific than the backbone counterparts. Finally, sequence specific assignments are obtained by a Monte Carlo simulated annealing protocol, which try to find an optimal fit among all plausible residue type assignments to the primary amino acid sequence and the 3D distribution of the protons.
The algorithm was tested for two small proteins, GAL (32 residues, predominantly α helical fold) and BPTI (58 residues, α/β twist fold). using experimental chemical shifts and synthetic (unambiguous) NOE restraints for all inter-proton distances shorter then 4Å. The 3D proton structures were within less than 2Å RMSD from the previously determined covalent 3D fine structures. The procedure fails when the distance restraints data set contain less than 70-80% of the theoretically obtainable data, and has no facilities to deal with overlap among ^{1}H^N-^{15}N

or ^{1}H-^{13}C chemical shift pairs.

1.2.2.2 The Atkinson-Saudek Method

Atkinson and Saudek extented the idea of direct conversion of NMR data to a (unconnected) structure by inclusion of information from through bond spectra [5]. By doing so, cross peaks from heteronuclear scalar coupling experiments are (beside those arising from NOESY spectra) also converted into distance restraints. For example, from the $^{1}H^{N}$-^{15}N HSQC yields a distance equal to the H-N bond length between hydrogen and nitrogen atoms ($\approx$ 1Å), the HN(CO)CA spectrum give rise for distances of 2.4Å and 2.5Å between backbone hydrogen H_i^N and nitrogen N_i to backbone C_i^α atoms, respectively, the HNCA spectrum yields distances of 1.5Å and 2.2Å between the same H_i^N/N_i pair and a second C_{i-1}^α atom (preceding in sequence). Similar treatment employing 2D $^{1}H^{N}$-^{13}C HSQC, 3D HNCO, 3D HN(CA)CO, 3D HNCACB, 3D CBCA(CO)NH, 3D HNHA and 3D HNHB spectra results in a extensive set of distances which can be inferred and converted into distance restraints for structure calculation:
C_i^α-C_i, C_{i-1}^α-C_i^α, C_{i-1}-C_i, C_{i-1}^α-C_i, C_{i-1}-C_i^α, C_i^α-C_i^β, C_{i-1}^β-C_i^α, C_{i-1}^α-C_i^β, C_i^β-C_i, C_{i-1}^β-C_i, C_{i-1}-C_i^β, C_{i-1}^β-C_i^β, C_i^β-H_i^α, C_i-H_i^α, C'_{i-1}-H_i^α, C_{i-1}^β-H_i^α, C_i-H_i^α, H_i^α-H_i^β, C_i^α-H_i^β, C_i-H_i^β, C_{i-1}^α-H_i^β, C_{i-1}^β-H_i^β, C_{i-1}-H_i^β.
Again, each atom therein is labeled with its chemical shift and the chemical shifts of the attached neighbour atoms.
The upper and lower bonds of the distance restraints were set according to the minimum and maximum values possible in model polypeptides (with a tolerance of 0.1Å).
3D structures of unconnected atoms are calculated applying a standard simulated annealing (SA) protocol, in which only a potential U_{NOE} for this set of distance restraints and a van der Waals term U_{vdW} were included only. The calculated structures contain only atoms observable by NMR. Breaks in the backbone chain occur at proline residues and unobserved backbone amide protons.
The authors took not shift degeneracy explicitly into account, they just demonstrated the principle performance of their approach. Although overlap of $^{1}H^{N}$ amide proton and ^{15}N heteronuclear chemical shift pairs can be readily identified by examining triple resonance spectra during input data preparation, overlap of H^α, C^α, and c^β, as well as the uncertainties of their cross peak positions are potential sources of data misinterpretation, which in turn could lead to extensive distortions in the calculated structures.

The procedure was tested using synthetic data for the protein ubiquitin, for which a high resolution crystal structure and full backbone and side chain ^{1}H, ^{13}C and ^{15}N chemical shift assignments are available. The lowest energy structures were found to closely resemble that of the crystal structure of ubiquitin, also in the case were 10% of the input data was randomly removed to account for incomplete data sets.

1.2.2.3 The CLOUDS method

In contrast to the presented direct methods mentioned above, the clouds protocol [50] was tested by using experimental data rather than simulated one. NOESY based inter-proton distances are obtained via relaxation matrix analysis (assuming isotropic tumbling and a single rotational correlation time τ_c) of unassigned, but unambiguous, experimental NOEs. Then proton distributions ('clouds') are calculated via a *md*-SA protocol, in which only a attractive potential U_{NOE} for the interproton distance restraints and a repulsive van der Waal potential term U_{vdW} were active. A second stage of the structure calculation include additional annealing by incorporation of *anti*-distance restraints (ATDRs) to bias the proton distribution compatible with the absence of specific NOEs. The implementation of ATDRs were found to significantly improve the convergence of the proton distributions. Both, the evaluation of a self-consistent NOE distance matrix (followed by converting them to distance restraints) and the application of ATDRs alleviate problems due to experimental NOE sparseness. Because ATDRs were set active only at the end of the MD simulation (at lower temperature), they were reported to do not influence the experimental NOE restraints. After filtering by reference to the proton distribution closest to the mean, a minimal dispersion proton density is obtained, whose interproton distances together with the individual atomic chemical shifts serves as experimental input of a subsequent Bayesian approach (for example see [11]) for the sequential assignment. In that analysis, the probabilities of sequential connectivity hypotheses are derived from likelihoods of H^N-H^H, H^N-H^α, and H^α-H^α interatomic distances as well as proton chemical shifts. Both sets of likelihoods are extracted from empirical databases. The empirical distance probability distributions were generated from a set of eight published high-quality protein structures solved via NMR and x-ray crystallography. The chemical shift PDs were extracted from BMRB database (BioMagResBank, http://www.bmrb.wisc.edu/). Once the protein backbone amide H^N and H^α atoms in the proton densities was assigned, the assignment of the side chain protons to the remainder proton densities (as dictated by the primary sequence) was performed in a similar

manner. Depending on the regional quality of the proton structure, even stereospecific assignments of the H^{β} protons could be feasible, because of the distinct empirical intra-residual distances between H^{β}_{proR} and H^{β}_{proS} to both, backbone H^{N} and H^{α} protons, respectively.
Into the assigned proton densities, the covalent geometry of the protein is fitted via a *md* protocol including potential energy terms for bonds, angles, torsions, and a repulsive-only non bonded term. For that purpose, the calculated ensemble of atomic distributions were digitized on a 0.5Å cubic grid, and smoothed with a Gaussian function to obtain a spatial 'continuous' proton density, which was converted to a pseudoenergy target function E_{pseudo} applying Boltzmann statistic. E_{pseudo} was combined with a standard protein forcefield ($PARALLHDG.PRO$ using xplor-NIH package [164]) for *md* embedding calculation. Smoothing of the original spatial proton densities facilitates a more proper sampling of conformational space and decreases ruggedness of energy landscape. Finally, the covalent structures were refined using explicit solvent. The *col2* domain of human matrix metalloprotease-2 (60 residues) and the *kringle2* domain of human plasminogen (83 residues) serves as test set for evaluation the procedure. Excellent unambiguous identified homonuclear NOEs list were available from previous, conventional structure analysis. The obtained structures deviate by 1.0 to 1.4Å RMSD for the backbone heavy atoms and 1.5 to 2.1Å RMSD for all heavy atoms, indicating the feasibility of successfully generating NMR structures from experimental NOESY data only. However, it's not a straightforward effort to generate a unambiguous NOE list of such completeness and quality used for these test calculations as input when starting a new structure determination.

1.2.3 Motivation

A part of this work concerned with the development and testing of a simple algorithm for rapid structure validation starting with a set minimum set of (experimentally derived) unambiguous NOEs related to backbone amide H^{N}, methyl groups, and optionally aromatic sidechain protons, without apriory knowledge of their sequential chemical shift assignments. As described in chapter 2, such minimal extent of unambiguous NOEs could generally easily be inferred even for larger sized proteins from a pair of 3D or 4D heteronuclear edited NOESY experiments by applying selective labeling. The benefit and the limitations of this structure generation approach were extensively evaluated using both, experimental and a idealized input data sets for several proteins of different folds.

Chapter 2

Methods

2.1 Protein NMR spectroscopy

2.1.1 Introduction

A detailed understanding of the function of biological macromolecules requires knowledge of their three dimensional structure. From NMR spectroscopy structural and dynamic information for proteins and protein-ligand complexes can be inferred. NMR provide useful information for generating a model basing on experimentally derived through bond interactions (via scalar coupling), through space interactions (per dipolar coupling) and chemical environments (from chemical shift). Furthermore, dynamic properties of proteins or domains can be characterized from measurement of relaxation data. Modern isotope labeling techniques and sensitive NMR experiments (e.g. employing TROSY method [138]), but also refined spectrometer equipment such as pulsed field gradients and cryo probes, facilitate structural and dynamical investigations of proteins up to molecular masses of about 30 kD.

The classical process of a protein conformation investigation using NMR spectroscopic methods involve:

- Preparation of a highly purified protein sample
- Acquisition of NMR spectra and processing
- Correlation of the spectra and generating of spin systems
- Sequential chemical shift and NOE assignment
- Extraction of conformational restraints

- Calculation of 3D structure models

Information from various stages in the process can be fed back to resolve ambiguities and correct errors at earlier steps. The used sample must be stable enough to permit acquisition for the whole set of required NMR spectra.

2.1.2 NMR tools and building blocks

The NMR experiments for chemical shift assignment and extraction of NOE connectivities are composed of several basic building blocks. Uniform or specific ^{13}c,^{15}N labeled samples with or without fractional deuteration allow extraction of NMR data for structure determination and dynamic studies for proteins up to 30 kD and more. Pulse sequences of multidimensional NMR experiments utilize the relatively large one and two bond couplings between pairs of heteronuclei and between heteronuclei and their directly attached protons for efficient coherence transfer. Spectral separation of the carbonyl C', aliphatic, and aromatic ^{13}C chemical shifts facilitate the selective manipulation of these resonances, allow handling them in different '^{13}C channels'. Appropriate phase cycles and the implementation of pulsed field gradients enables selection of the signal of interest, the elimination of artifacts, and the suppression of the water signal. Some of the NMR experiment building blocks and principles used in this work are listed below.

2.1.2.1 Water suppression techniques

The concentration of protons in water is approximately 110 mM compared with the typical concentration of 1-2 mM of the investigated protein. Consequently, the equilibrium magnetization of the water protons is $\approx 10^4$ to 10^5 times greater than the equilibrium magnetization of a single proton resonance in the protein, which causes overflow in the analog to digital converter and distortion of the protein signals by the large broad solvent peak. Further complications arise from the radiation damping of the solvent signal. Application of deuterium oxide (D_2O) as solvent (instead of H_2O) reduces the resonances from exchangeable protons of the protein as well.

Presaturation

A classical solvent-suppression method is presaturation of the solvent signal during the recycle delay using a soft rf field, which works very effective. However, protein signals (mainly H^{alpha} protons) that resonate very closely to the water signal may be partially saturated. Furthermore,

resonances of exchangeable protons (as amide protons ^{H}N) will also be effected via saturation transfer.

Jump-Return technique

In this method [134], the final read pulse of an experiment (generating detectable transversal magnetization) is replaced by the pulse element 90_x-τ-90_{-x}. The carrier frequency of the rf field is placed on the water resonance and the delay according $\tau = 1/(4\Delta\nu_{max})$, whereas ν_{max} is the offset from the carrier at which excitation is maximized (typically the amide proton region). No transverse magnetization is generated for protons with zero offset (as the solvent). The detected signals for coherences that are not resonant with the rf carrier are modulated by a factor $\sin(\Omega\tau)$ (with Ω the frequence difference between the rf carrier and a individual spin).

Spin lock purge pulses and pulsed field gradients

In the simplest approaches [169], spin lock purge (SL) pulses are applied (for a duration of about 1-2 ms) to dephase the magnetization orthogonal orientated to it, while *spin locking* the collinear magnetization of interest. In a simple example of water suppression, a selective 90°_{x} pulse is used to rotate the water magnetization into the transversal plane. A non-selective 90°_{x} hard pulse flip the solvent magnetization back to the longitudinal axis (described by operator S_z), while bringing the solute magnetization transversal ($-I_y$). A subsequent spin lock purge pulses of phase y lock the transversal solute magnetization along the y-axis, whereas the longitudinal solvent magnetization is dephased rapidly by the rf B_1 inhomogeneity (which is very large compared to static B_0).

Beside coherence selection and artifact elimination, pulsed field gradients are also used in water suppression. In the *WATERGATE* technique [139], after an initial non-selective 90° pulse (for example of phase x) a strong gradient pulse dephase both, the solvent and the solute magnetization, in the middle of a delay Δ. Subsequent a pulse element $90^{\circ}_{-y,solvent}$-180°_{y}-$90^{\circ}_{-y,solvent}$ is applied, in which $90^{\circ}_{-y,solvent}$ denotes a soft pulse selective to the solvent. The non-selective 180°_{y} pulse invert the coherence order of the solute, whereas the combination of the selective $90^{\circ}_{-y,solvent}$ pulses and the non-selective 180°_{y} pulse let the coherence order of the solvent unchanged. A second gradient pulse (of the same sign and amplitude as the first) rephases the magnetization of the solute to form a gradient echo, whereas the magnetization of the solvent is being further dephased.

Another technique is denoted as *'water flip back'* method [32], in which a selective $90^{\circ}_{-x,solvent}$ is inserted at the begin of the *WATERGATE* sequence. Consequently, the solute magnetization is described by an operator $-I_y$ and the solvent by longitudinal S_z. Again, the pulsed field gradient

flanking the pulse element $90^\circ_{-y,solvent}$-180°_y-$90^\circ_{-y,solvent}$ results in a gradient echo for the solute magnetization. In contrast to above, the solvent magnetization is aligned along the longitudinal z-axis, thus never saturated in this experiment. Any transverse solvent magnetization resulting from non-ideal selective 90° and/or non-selective 180° pulses is dephased by the field gradient pulse. Because in praxis the main fraction of the water magnetization is maintained along the z-axis, a remarkable increased sensitivity are obtained for exchangeable protein protons due to the reduced saturation transfer.

2.1.2.2 INEPT sequence

An *Insensitive Nuclei Enhanced by Polarization Transfer* (INEPT) sequence [122] is applied in order to transfer magnetization from a sensitive nucleus I (typically protons) with a high gyromagnetic ratio to a less sensitive nucleus S offering a lower gyromagnetic ratio (heteronucleus ^{13}C or ^{15}N) via their scalar coupling interaction. By doing so, the detectable signal of the heteronucleus ^{13}C or ^{15}N is increased by a factor of $\gamma_{^1H}\gamma_{^{13}Cor^{15}N}$. An INEPT pulse block is written as:

I spin:	90°_x	- t -	180°_x	- t -	90°_y
S spin:			180°_x		90°_x

The pulse element -t-180°_x(I,S)-t- is a heteronuclear spin echo, in which the chemical shift is refocused and the heteronuclear coupling J_{IS} evolves fully as:

$$I_z \xrightarrow{\frac{\pi}{2}I_x - t - \pi(I_x+S_x) - t} I_y \cos(2\pi J_{IS} t) - 2I_x S_z \sin(2\pi J_{IS} t) \tag{2.1}$$

The final 90°_x(I) and 90°_x(S) pulse pair generates:

$$\begin{aligned} &I_y \cos(2\pi J_{IS} t) - 2I_x S_z \sin(2\pi J_{IS} t) \xrightarrow{\frac{\pi}{2}(I_x+S_x)} \\ &\qquad I_y \cos(2\pi J_{IS} t) - 2I_z S_y \sin(2\pi J_{IS} t) \end{aligned} \tag{2.2}$$

If the delay t is set to $1/4JIS$, then the final signal is given by $-2I_zS_y$.

In the *refocused* INEPT version an additional heteronuclear spin echo (-t-180°_x(I,S)-t; with $t = 1/4JIS$) is following in order to convert the anti-phase coherence $-2I_zS_y$ into an observable in-phase coherence S_x.

2.1.2.3 TOtal correlation spectroscopy (TOCSY) coherence transfer through bonds

By dint of *isotropic mixing sequences* [98], which permit more step through bond coherence transfer via strong scalar coupling interaction, potentially all spins within a scalar coupled network can be correlated to each other. Therefore experiments containing such sequences are denoted as TOtal correlation spectroscopy (TOCSY). Such (composite) pulse sequences facilitates efficient magnetization transfer via fulfilling the Hartmann-Hahn condition (in which all chemical shift and rf pulse terms in the Hamiltonian governing the spin system have identical values) by generating an average Hamilton for the coupled spinsystem during a mixing time period τ_m, which consists only of a scalar coupling term. The most common mixing pulse element is the DIPSI-2 sequence offering significant larger coherence transfer efficiencies than other successfully employed sequences of the MLEV and WALTZ family of pulse sequences.
In a TOCSY experiment, the spins are frequency labeled during t_1 and returned to longitudinal z-axis using the 90°-t_1-90° element. During the subsequent isotropic mixing time τ_m magnetization can be transfered potentially between all spin of the scalar coupled network prior application of a read pulse and acquisition. In practice, the duration of τ_m is chosen to obtain maximum information about the spinsystem while minimizing cross peaks intensity loses due to relaxation.

2.1.2.4 Through space coherence transfer - the NOESY experiment

Coherence can also be transfered through space within a dipolar coupled spin network [126]. Initially, a 90°_x-t_1-90°_x period frequency labels the spins and returns the magnetization back to z-axis. Omitting scalar coupling one obtains for a spin I_1:

$$I_{1z} \xrightarrow{\frac{\pi}{2}I_x - t_1 - \frac{\pi}{2}(I_x)} -I_{1z}\cos(\Omega_1 t_1) + I_{1x}\sin(\Omega_1 t_1) \tag{2.3}$$

The second term I_x is suppressed by phase cycling or field gradient pulse. The inverted populations along the I_{1z} spin transitions induces a NOE to a dipolar coupled spin I_k during a mixing time τ_m:

$$-I_{1z}\cos(\Omega_1 t_1) \xrightarrow{\tau_m} -\sum_{k=1}^{N} I_{kz} a_{1k}(\tau_m)\cos(\Omega_1 t_1) \tag{2.4}$$

whereas $a_{11}(\tau_m)$ are related to the fraction of the original magnetization remaining on spin I_1, and $a_{1k}(\tau_m)$ the transfered magnetization to a dipolar coupled spin I_k. Particularly, $a_{1k}(\tau_m)$ is

the (1,k)th element of the matrix exponential $\exp(-\mathrm{R}\tau_m)$, with R the matrix of the auto relaxation rate ρ_i and cross-relaxation rate σ_{ij} constants.

After a final 90° read pulse the final spectrum contain I_1 diagonal peaks and I_k cross peaks (with $k > 1$). Cross peaks in a NOESY spectrum can also arise from slow chemical exchange (on the chemical shift time scale), if the exchange rate of a particular nucleus is not slow compared to τ_m. For molecules as big as proteins, such chemical exchange peaks offer the same sign as NOE cross peaks, which make discrimination difficult.

2.1.2.5 Sensitivity improvement by 'preservation of equivalent pathways' (PEP)

This method facilitates sensitivity enhancement by converting both orthogonal transverse magnetizations components, which can be obtained in the indirect dimensions, into detectable magnetization. The concept, developed by Rance and Co-workers [23], provide sensitivity improvement up to $\sqrt{2}$ relative to the conventional experiment, in which only the half of the initial spin polarization does contribute to the final signal. For visualization, the PEP variant of an 2D HSQC experiment is described. After an initial INEPT sequence, which transfers in phase magnetization presented by I_y to anti phase coherence -$I_z S_y$, and frequency labeling during evolution time t_1 the two orthogonal terms $2I_z S_y \cos(\Omega_s t_1)$ and $-2I_z S_x \cos(\Omega_s t_1)$ are obtained. In the conventional HSQC, a subsequent reverse INEPT element 90°(I_x, S_x)-τ-180°(I_x, S_x)-τ would refocus the first term only (to $-I_x \cos(\Omega_S t_1)$,giving rise to an amplitude modulated signal), whereas the second remain as non-observable multiple quantum coherence ($-2I_y S_x \sin \Omega_S t_1$). The situation is different if an additional spin echo is attached to the reverse INEPT sequence as follows:

$$2I_z S_y \cos(\Omega_s t_1) - 2I_z S_x \cos(\Omega_s t_1) \xrightarrow{\frac{\pi}{2} I_x \frac{\pi}{2} S_x - \tau - \pi(I_y + S_x) - \tau - \frac{\pi}{2} I_y \frac{\pi}{2} S_y - \tau - \pi(I_y + S_y) - \tau - \frac{\pi}{2} I_x} -I_y \cos(\Omega_s t_1) - I_x \sin(\Omega_s t_1) \qquad (2.5)$$

Both terms in equation 2.5 are related to observable orthogonal *in-phase* I spin magnetization and represent a superposed (phase modulated) signal with a 90° phase difference in $\omega 1$ and $\omega 1$ dimensions (resulting in phase-twisted lineshapes). Refocusing of both signal components is only possible for an IS spin system, homonuclear coupling of I to another spin would lead to unobservable multi quantum coherence. Therefore, this extended reverse polarization transfer implements a spin multiplicity filter. To obtain pure absorptive lineshapes, a second experiment is performed in which the phase of the 90°(S_x) pulse immediately performed after the t_1

evolution period is inverted to 90°(S_{-x})

$$2I_zS_y\cos(\Omega_s t_1) - 2I_zS_x\cos(\Omega_s t_1) \xrightarrow{\frac{\pi}{2}I_x\,\frac{\pi}{2}S_{-x}-\tau-\pi(I_y+S_x)-\tau-\frac{\pi}{2}I_y\,\frac{\pi}{2}S_y-\tau-\pi(I_y+S_y)-\tau-\frac{\pi}{2}I_x} I_y\cos(\Omega_s t_1) - I_x\sin(\Omega_s t_1) \quad (2.6)$$

Addition of both data sets in equation 2.5 and 2.6 results in a single observable term $-2I_x\sin(\Omega_s t_1$ (note that this signal only arise from a IS spin system), whereas subtraction gives $-2I_y\cos(\Omega_s t_1)$ (this data set also contains signals from I_nS spin systems). By doing so, both orthogonal terms in 2.5 and 2.6 are separated and each of the newly generated data sets can be transformed and phased to pure absorptive peak shapes in both dimensions. The two spectra are added to obtain the final sensitivity enhanced spectrum.

2.1.2.6 Isotope filtering

In most of heteronuclear experiments one is only interested to select signals of protons directly attached to a heteronucleus (in order to simplify the spectrum). In the example above this is accomplished by simultaneously inverting the phase of the 90°(S_x) pulse immediately performed after the t_1 evolution (where the terms $2I_zS_y\cos(\Omega_s t_1)$ and $-2I_zS_y\cos(\Omega_s t_1)$ are existent from the scalar coupled I spins) period and the receiver. Signals of I spins, which are not scalar coupled to a S spins are unaffected by the inversion of the 90°(S_{-x}) pulse and are canceled by inversion of the receiver phase (in progress of the signal accumulation).

2.1.3 Heteronuclear experiments for chemical shift assignment

Triple resonance experiments (listed below) introduced by Bax and Co-workers are exploited since the early 90ies for sequential chemical shift assignment and correlate backbone $^1H^N$, ^{15}N, $^{13}C^\alpha$, $^{13}C^\beta$, and C' spins by using one and two bond couplings. A second generation of them provide also connectivities from the whole sidechain spinsystem to the backbone amide via TOCSY through bond transfer. These experiments were subsequent refined, e.g. by introducing pulsed field gradients for coherence selection and artifacts suppression associated with 180° pulses, or by using the PEP technique for sensitivity enhancement. A short overview over some variants of these experiments used for chemical shift assignment in this work are listed after description of the 2D HSQC experiment below.

2.1.3.1 2D ^{1}H-^{15}N sensitivity enhanced HSQC

This experiment correlates the backbone amid proton to their directly attached nitrogen atom. Via an initial INEPT transfer pulse segment magnetization is transfered from the sensitive amide proton to the nitrogen as heteronuclear single quantum coherence (HSQC) $-H_z^N N_y$. During the subsequent t_1 evolution time this coherence is frequency labeling while heteronuclear coupling is refocused by a 180°(H^N) pulse. A reverse INEPT sequence transfers the magnetization back to the amide protons and subsequent a sensitivity enhancement using the PEP technique is applied prior detection period t_2, at which decoupling is achieved using the GARP-1 decoupling sequence. Water suppression is performed by incorporation of a spin lock purge pulse into the initial INEPT sequence after the end of the spin echo and prior the final 90°(H_y^N, N_x) pulse pair (where the magnetization of the solvent and the solute are orthogonal to each other).
Additional cross peaks other than from backbone amide functions are potentially observable for glutamine and asparagine side chain amide groups, arginine side chain H^{N_ϵ}/N^ϵ and H^{N_ζ}/N^ζ, and tryptophan indole H^{N_ϵ}/N^ϵ functions. Due to rapid exchange with solvent water, signals for histidine imidazole H^{N_δ}/N^δ or $H^{N_{\epsilon 2}}$/$N^{\epsilon 2}$ are commonly not observed.

2.1.3.2 3D HNCA

This experiment [124] correlates the backbone amide ^{1}H^N and ^{15}N chemical shifts with the intra-residue C$_i^\alpha$ shift by employing the small one bond J_{NC^α} coupling (about 11 Hz) and the C$_{i-1}^\alpha$ carbon chemical shift of the preceding residue via inter-residue two bond J_{NC^α} coupling (about 9 Hz, exact values depending on the conformation state).
By an initial INEPT sequence magnetization is transfered from backbone amide proton to the bounded nitrogen. A second INEPT step generates coherence transfer to intra-residue ^{13}C$_i^\alpha$ and preceding ^{13}C$_{i-1}^\alpha$ carbon spins. During subsequent ^{13}C frequency labeling in t_2, heteronuclear J_{NC^α} and $J_{C^\alpha C'}$ coupling evolution is refocused by 180°(C') and 180°(N) pulses. After an 90°(C^{alpha})/90°(N) pulse pair a constant time period follows, in which ^{15}N frequency labeling during t_1 and J_{NC^α} coupling (during CT period T) development occurs simultaneously. Finally, a PEP reverse INEPT transfers the coherence back into observable amide H^N magnetization. Evolution of the one bond $J_{H^N N}$ coupling between the first INEPT plus a delay τ' and a second delay τ' preceding the final PEP reverse INEPT step is prevented by applying synchronous broadband ^{1}H decoupling using WALTZ-16 sequence, whereas τ' is an appropriate delay $(1/2J_{H^N})$ for converting anti-phase $-H_z^N N_y$ to in-phase ^{15}N magnetization in respect to

the attached amide proton.

The magnetization transfer is given by:

$$^1H_i^N \xrightarrow{^1J_{NH^N}} {}^{15}N_i \xrightarrow{^1J_{NC^\alpha},^2J_{NC^\alpha}} C^\alpha_{i,i-1}(t_1) \rightarrow {}^{15}N(t_2) \rightarrow {}^1H_i^N(t_3)$$

2.1.3.3 3D HN(CO)CA

The 3D HN(CO)CA [124] provides sequential correlation between the amide ^{1}H^N and nitrogen ^{15}N chemical shift to the ^{13}C$^\alpha$ chemical shift of the preceding residue by coherence transfer via the intervening ^{13}C' spin. Nitrogen ^{15}N frequency labeling is performed in a constant time period immediately after an INEPT step starting from backbone amide. By a 90°(C')/90°(N) pulse pair magnetization is transfered to the carbonyl spin, which is subsequent correlated with the attached ^{13}C$^\alpha$ spin in a HMQC like manner. Here ^{13}C$^\alpha$ frequency labeling occurs during t_2. As before ^{1}H decoupling is facilitated via WALTZ-16 sequence between the initial INEPT and the concluding PEP reverse INEPT step.

The magnetization pathway is presented by:

$$^1H_i^N \rightarrow^{15} N_i \rightarrow^{13} C'_{i-1} \rightarrow^{13} C^\alpha_{i-1}(t_1) \rightarrow^{13} C'_{i-1} \rightarrow^{15} N(t_2) \rightarrow^1 H_i^N(t_3)$$

2.1.3.4 3D HNCACB

The 3D HNCACB experiment [195] correlates the backbone amide ^{1}H^N and ^{15}N chemical shifts with the intra-residue C$^\alpha_i$ and C$^\beta_i$ shifts and the C$^\alpha_{i-1}$ and C$^\beta_{i-1}$ carbon chemical shifts.

The pulse sequence of the experiment is very similar to that of the 3D HNCA experiment. After the standard INEPT transfer from ^{1}H^N to ^{15}N and refocusing the anti-phase ^{15}N magnetization to inphase in respect to bounded ^{1}H^N spin, proton decoupling is performed as before (using WALTZ-16). A second INEPT step transfers coherence from ^{15}N to scalar coupled C$^\alpha_{i,i-1}$ spins to give $-N_zC^\alpha_y$ magnetization. Now the sequence differ: In a subsequent period (equal to $1/4J_{C^\alpha C^\beta}$; with $J_{C^\alpha C^\beta} \approx$ 35 Hz) C$^\alpha C^\beta$ coupling is evolved while refocusing the ^{13}C$^\alpha$ and ^{13}C$^\beta$ chemical shifts by an 180°(C$_{aliphatic}$) pulse. Then, during t_1 both shifts are developed for ^{13}C frequency labeling, whereas nitrogen ^{15}N and carbonyl ^{13}C' are decoupled by a 180°(N) and homonuclear SEDUCE-1 decoupling, respectively. In contrast, the $J_{C^\alpha C^\beta}$ and $J_{C^\beta C^\gamma}$ couplings remain active during t_1, causing slight line broadening in the ω_1 dimension. In a second $1/4J_{C^\alpha C^\beta}$ period magnetization is focused back to ^{13}C$^\alpha$ spins. A pair of 90°(C$_{aliphatic}$)/90°(N)

pulses transfers to magnetization back to ^{15}N. During subsequent $\Delta - t_2/2$-180°(N,C$_aliphatic$)-$\Delta + t_2/2$ ^{15}N (while C' and ^{1}H is still active) is encoded in t_2 in a constant-time manner. A final reverse INEPT step generates amide proton magnetization for detection in t_3.
In summary, the magnetization follows the pathway:

$$^1H_i^N \xrightarrow{^1J_{NHN}} {}^{15}N_i \xrightarrow{^1J_{NC^\alpha},{}^2J_{NC^\alpha}} C^\alpha_{i,i-1} \xrightarrow{^1J_{C^\alpha C^\beta}} C^\alpha_{i,i-1}/C^\beta_{i,i-1}(t_1) \xrightarrow{^1J_{NC^\alpha},{}^2J_{NC^\alpha}} C^\alpha_{i,i-1} \rightarrow {}^{15}N(t_2) \rightarrow {}^1H_i^N(t_3)$$

2.1.3.5 3D CBCA(CO)NH

This experiment [52] provides connectivity information about the backbone amide ^{1}H^N and ^{15}N chemical shifts of residue i to the ^{13}C$^\alpha$ and ^{13}C$^\beta$ resonances of the preceding residue $i-1$ solely.
The pulse scheme starts with an INEPT transfer from the aliphatic ^{1}H$^\alpha$ and ^{1}H$^\beta$ protons to bounded ^{13}C$^\alpha$ and ^{13}C$^\beta$ carbon spins. All ^{13}C$^{\alpha,\beta}$ pulses are applied near the center of the ^{13}C$^\alpha$ and ^{13}C$^\beta$ chemical shift ranges in order to maximize their exication, whereas the power of the ^{13}C$^{\alpha,\beta}$ pulses are adjusted in order to minimize their effect on the carbonyl C' spins. Then, a constant time period sequence $t_1/2$-T-180°(C$^{\alpha,\beta}$)-T-$t_1/2$ is following, at which ^{13}C$^\alpha$/^{13}C$^\beta$ frequency labeling is performed. Scalar coupling to carbonyl C' spin is removed by a 180°(C') pulse after time $t_1/2$. To eliminate its influence on the ^{13}C$^\alpha$/^{13}C$^\beta$ frequency (arising from off resonance effects) the same 180°(C') pulse is applied at the end of the constant time period T. A subsequent 90°(C$^{\alpha,\beta}$) pulse convert (anti-phase) C$^\beta$ magnetization to (anti-phase) C$^\alpha$ magnetization. Afterwards, a second INEPT step transfers magnetization from ^{13}C$^\alpha$ to the adjacent ^{13}C' spins employing the large $J_{C'C^\alpha}$ coupling (about 55 Hz). A third INEPT transfer generates backbone ^{15}N magnetization, followed by an second constant time period, in which ^{15}N frequency labeling occurs. Then, a final reverse INEPT step generates backbone amide magnetization for detection in t_3.
In total, the magnetization pathway can be visualized as follows:

$$^1H^\alpha_{i-1}/{}^1H^\beta_{i-1} \xrightarrow{^1J_{H^\alpha C^\alpha},J_{H^\beta C^\beta}} C^\alpha_{i-1}/C^\beta_{i-1}(t_1) \xrightarrow{^1J_{C'C^\alpha}} {}^{13}C'_{i-1} \xrightarrow{^1J_{C'N}} {}^{15}N_i(t_2) \rightarrow {}^1H_i^N(t_3)$$

2.1.3.6 3D HNCO

The 3D HNCO experiment [124] facilitate the connection between the backbone amide ^{1}H^N and ^{15}N chemical shifts of residue i to the ^{13}C' carbon shift of the preceding residue $i-1$ via $J_{NC'}$ coupling (about 15 Hz).

By interchanging the C^{α} and C' pulses the 3D HNCO pulse sequence employed is the same as for the 3D HNCA experiment.

The magnetization transfer is presented by:

$$^{1}H_{i}^{N} \xrightarrow{^{1}J_{NH^{N}}} {}^{15}N_{i} \xrightarrow{^{1}J_{NC'}} C'(t_{1}) \rightarrow {}^{15}N(t_{2}) \rightarrow {}^{1}H_{i}^{N}(t_{3})$$

2.1.3.7 3D HN(CA)CO

The 3D HN(CA)CO [124] connects the chemical shifts of the intra-residue $^{1}H_{i}^{N}$, $^{15}N_{i}$, and $^{13}C_{i}^{'}$ spins via one bond $^{1}J_{NC^{\alpha}}$ ($\approx$ 11 Hz) and $^{1}J_{C'C^{\alpha}}$ ($\approx$ 55 Hz) couplings. In addition, correlation the carbonyl spin of the preceding residue is potentially provided using inter-residue two bond $^{2}J_{NC^{\alpha}}$ coupling. However, because of the rather small $^{2}J_{NC^{\alpha}}$ coupling constant value (about 7 Hz or less), this connection feature hardly depends on the relaxation properties of the protein. In a folded protein, normally only a fraction can be observed.

Starting from amide proton $^{1}H^{N}$ coherence, magnetization is transfered to the $^{13}C^{\alpha}$ spins by using two INEPT steps. The subsequent $^{13}C^{\alpha} \rightarrow {}^{13}$C' transfer is implemented as a HSQC, whereas during ^{13}frequency labeling the $J_{C'C^{\alpha}}$ and $J_{NC^{\alpha}}$ couplings are refocused by a 180°(C^{α},N) pulse pair in the middle of the evolution time t_{2}. After a 90°(C^{α},N) pulse pair ^{15}N frequency labeling is performed in a constant time period. A final PEP reverse INEPT sequence generates amide proton magnetization again, which is detected during t_{3}. Where necessary, broadband ^{1}H decoupling is achieved using DIPSI-2 sequence.

The magnetization transfer pathway is given by:

$$^{1}H_{i}^{N} \xrightarrow{^{1}J_{NH^{N}}} {}^{15}N_{i} \xrightarrow{^{1}J_{NC^{\alpha}},^{2}J_{NC^{\alpha}}} C_{i,i-1}^{\alpha} \xrightarrow{^{1}J_{C'C^{\alpha}}} {}^{13}C_{i,i-1}(t_{1}) \xrightarrow{^{1}J_{C'N}} {}^{15}N_{i}(t_{2}) \rightarrow {}^{1}H_{i}^{N}(t_{3})$$

2.1.3.8 3D HNN

The 3D HNN experiment [136] sequentially correlates the intra-residue $^{1}H^{N}$, ^{15}N chemical shifts and the ^{15}N chemical shift to each other by employing the rather small $^{1}J_{NC^{\alpha}}$ ($\approx$ 11 Hz) and $^{2}J_{NC^{\alpha}}$ ($\approx$ 7 Hz) couplings. Thus, this experiment is only feasible for polypeptides or domains showing a long transverse T_{2} time values, as present in flexible and intrinsic unfolded proteins. However, especially in such cases this experiment (in context with the 3D HN(C)N, see below) is highly valuable for backbone assignment because of the typically poor ^{1}H and ^{13}C chemical shift dispersion occurring in these structures.

The HNCA based pulse sequence starts with an INEPT transfer from backbone amide $^1H^N$ to the bounded ^{15}N followed by a constant time period including evolvement of the $^1J_{NC^\alpha}$ and $^2J_{NC^\alpha}$ couplings as well as ^{15}N frequency labeling during t_1. Subsequent a 90°(C^α,N) pulse pair transfers magnetization from $^{15}N_i$ to $^{13}C^\alpha_{i,i-1}$ spins. During a following delay τ_{CN} $^{13}C_{i,i-1}$ spin anti-phase magnetization is generated in respect to their own ^{15}N nucleus via one bound $^1J_{NC^\alpha}$ and to neighboring ^{15}N spins per $^2J_{NC^\alpha}$ couplings, respectively. By another 90°(C^α,N) pulse pair magnetization is relayed back to the ^{15}N spins (of residue $i-1$, i, and $i+1$) followed by a second constant time period containing t_2. By a PEP reverse INEPT element magnetization is transfered to directly attached amide spins $^1H^N$, which are detected in t_3.
In short, the magnetization transfer through the pulse sequence is given by:

$$^1H_i^N \xrightarrow{^1J_{NH^N}} {}^{15}N_i(t_1) \xrightarrow{^1J_{NC^\alpha},^2J_{NC^\alpha}} C^\alpha_{i,i-1} \xrightarrow{^1J_{NC^\alpha},^2J_{NC^\alpha}} {}^{15}N_{i-1}(t_2),^{15}N_i(t_2),^{15}N_{i+1}(t_2) \xrightarrow{^1J_{NH^N}} {}^1H^N_{i-1}(t_3),^1H^N_i(t_3),^1H^N_{i+1}(t_3)$$

2.1.3.9 3D HN(C)N

This experiment [136] correlates the ^{15}N chemical shifts of residue i to the $^1H^N$ and ^{15}N chemical shifts of the preceding residues $i-1$ using a pulse sequence basing on the 3D HN(CO)CA experiment, at which (in contrast to the HNN experiment) magnetization is transfered via $JC'N$ and $JC'C^\alpha$ couplings to the $^{13}C^\alpha_{i-1}$ spins solely. The experiment is particular useful in connection with the HNN experiment.
The magnetization transfer is presented by:

$$^1H_i^N \xrightarrow{^1J_{NH^N}} {}^{15}N_i(t_1) \xrightarrow{^1J_{NC'}} C'_{i-1} \xrightarrow{^1J_{C'C^\alpha}} {}^{15}N_{i-1}(t_2),^{15}N_i(t_2) \xrightarrow{^1J_{NH^N}} {}^1H^N_{i-1}(t_3),^1H^N_i(t_3)$$

2.1.3.10 3D H(CCO)NH-TOCSY

The 3D H(CCO)NH-TOCSY experiment [51] correlate the aliphatic ^1Hi resonances directly with the backbone amide $^1H^N_{i+1}$ of the subsequent residue.
The pulse sequence employed is very similar to 3D CBCA(CO)NH. The main difference is an isotropic ^{13}C mixing sequence applied to transfer the magnetization from the aliphatic sidechain to $^{13}C^\alpha$ spin. After an initial INEPT transfer from the aliphatic ^{13}C magnetization is anti-phase with respect to their directly attached protons. In a subsequent constant time period T this magnetization is refocused, ^{13}C frequency labeling occurs in t_1 and the $J_{\alpha\beta}$ coupling evolves

fully during T. A short spin-lock ('trim') pulse follows, which dephases the anti-phase $C_y^\beta C_z^\alpha$ coherence that is not parallel to the effective field. The subsequent isotropic mixing sequence applied for time τ_{mix} transfers in-phase C_y^β to its neighbors within the ^{13}C carbon spin system. Because the ^{13}C chemical shifts are dispersed over a wide frequency range, an efficient broadband isotropic mixing scheme (as DIPSI-3 or FLOPSY) have to be employed to achieve the Hartmann Hahn condition. In two following INEPT steps magnetization is transfered via the carbonyl ^{13}C' to backbone ^{15}N spins. Subsequent ^{15}N frequency labeling occurs during t_2 of a constant time sequence fragment prior the magnetization is transfered to the amide protons by a finishing INEPT transfer. Between the end of the mixing time τ_{mix} and the last INEPT step ^{1}H broadband decoupling is performed using DIPSI-2 decoupling sequence.
The magnetization pathway through the sequence can be summarized as:

$$^1H_{i-1}^{aliphatic} \xrightarrow{^1J_{HC}} C_{i-1}^{aliphatic}(t_1) \xrightarrow{1.\tau_{mix}, 2.^1J_{C'C^\alpha}} C_{i-1}^{\alpha} \xrightarrow{^1J_{C'C^\alpha}} {}^{13}C'_{i-1} \xrightarrow{^1J_{C'N}} {}^{15}N_i(t_2) \;\rightarrow\; {}^1H_i^N(t_3)$$

2.1.3.11 3D C(CO)NH-TOCSY

The 3D HC(CO)NH-TOCSY experiment [51] provide the correlation the aliphatic ^{13}Ci resonances with the backbone amide ^{1}H$_{i+1}^N$ of the residue following in the primary sequence.
The pulse sequence of the 3D HC(CO)NH-TOCSY experiment is virtually identical to the 3D H(CCO)NH-TOCSY experiment, but here the ^{1}H spins are frequency labeled during t_1, which is inserted in a modified INEPT sequence. Originally, the evolution time t_1 and a normal INEPT sequence fragment were arranged consecutively (case A). In the applied pulse sequence, both were merged in a single block of the form:

^{1}H spin:	90°_x	- τ -	t_1	180°_x	- τ -	90°_y
^{13}C spin:			180°_x			90°_x

The 180°(^{13}C) pulse in the middle of t_1 ensures decoupling of the ^{1}H spins from the ^{13}C spins during this evolution period. Applying of only one 180°(^{13}C) pulse instead of two (as performed when evolution time t_1 and INEPT sequence are serialized in a subsequent manner) reduce artifacts from pulse imperfections. Furthermore, multiple quantum coherence is generated during t_1 which relaxes more slowly than single quantum coherence. Ignoring pulse artifacts and relaxation, the result at the end of the sequence element shown above is identical as obtained from the original version (case A).

The coherence pathway is given by:

$$^{1}H_{i-1}^{aliphatic}(t_1) \xrightarrow{^{1}J_{HC}} C_{i-1}^{aliphatic} \xrightarrow{1.\tau_{mix},2.^{1}J_{C'C\alpha}} C_{i-1}^{\alpha} \xrightarrow{^{1}J_{C'C\alpha}} {}^{13}C'_{i-1} \xrightarrow{^{1}J_{C'N}} {}^{15}N_i(t_2) \;\rightarrow\; {}^{1}H_i^{N}(t_3)$$

2.1.4 Measurement of three bond couplings

The measurement of the $^{3}J_{C'N}$ and $^{3}J_{C'C}$ provides conformational information about the side chain torsion χ_1 angle. The ^{13}CO and ^{15}N spin echo constant time HSQC experiment described below were used in this work for estimation of the χ_1 torsion angle of valine, isoleucine and threonine residues.

2.1.4.1 ^{13}CO spin echo constant time HSQC

The ^{13}CO spin echo constant time HSQC experiment [53] allows the determination of the coupling constant $^{3}J_{CC}$ between the backbone carbonyl ^{13}C' and the methyl carbons -$^{13}CH_3$ of valine, isoleucine ($C_{\gamma 2}$), and threonines. At the end of an initial INEPT step the aliphatic $^{13}C^{aliphatic}$ spin magnetization is anti-phase in respect to the aliphatic protons $^{1}H^{aliphatic}$. Subsequent, a constant time period T sequence fragment is following of the form:
-$t_1/2 + \delta$-180°($^{1}H^{aliphatic}$)-$T/2 - \delta$-180°($^{13}C^{aliphatic}$)-$T/2$-$t_1/2$-
The sequence permits the definition about the evolution of heteronuclear coupling $^{1}J_{CH}$ while frequency labeling occurs during t_1. For $\delta = 0$ $^{1}J_{CH}$ is active for a total time $+t_1/2 - T/2 + T/2 - t_1/2 = 0$, so $^{13}C^{aliphatic}$ magnetization remains anti-phase relative to $^{1}H^{aliphatic}$ spins. Homonuclear coupling J_{CC} is active for the entire constant time period T ($+t_1/2 + T/2 + T/2 - t_1/2 = T$), because the non-selective 180°($^{13}C^{aliphatic}$) pulse effect the coupled spins in the same way. Because homonuclear scalar coupling is independent of t_1 the ω_1 dimension offers no homonuclear multiplet structure, which enhances the resolution. However, the peak intensities I in ω_1 become a function of $^{1}J_{CC}$ (by a factor of $\cos(2\pi J_{CC}T)$ or $\sin(2\pi J_{CC}T)$ depending on the phase cycle). The one bond aliphatic coupling $^{1}J_{CC}$ ranges closely about 35 ($\pm$ 3) Hz, which enables the maximization of the peak intensities I as a function of T only. An useful feature becomes apparent when considering the contributions of the total transverse relaxation R_2^* in a constant time sequence fragment. The homogeneous relaxation rate R_2 active during total constant period T, whereas the inhomogeneous part of R_2^* is partly refocused by the 180°($^{13}C^{aliphatic}$) pulse and in total present only during $+t_1/2 + T/2 - (T/2 - t_1/2) = t_1$. Thus, the related time signal is proportional to $e^{-R_2 T} \cdot e^{-R_{inhom} t_1}$. The first (constant) term ($e^{-R_2 T}$) is a multiplicative

factor reducing the intensity of the peak signals, whereas the damping rate $e^{-R_{inhom}t_1}$ determines the line width in ω_1. Because in practice $R_{inhom}t_1 \ll 1$, the ω_1 line width obtained simply depends on the window function applied to apodize the FID.

Decoupling to the ^{15}N during T is achieved using a DIPSI-2 decoupling sequence. Evolution of the coupling to the carbonyl ^{15}C' spins is performed by a 180°(^{13}C') pulse after $t_1/2$. Aliphatic ^{13}C evolution caused by off-resonance effect of this pulse is refocused by another 180°(^{13}C') pulse at the end of the time period T. The sequence is finished by a reverse INPET sequence element.

The obtained signal acquired in that way is stored and denoted here as $S_{^3J_{CC'}-decoupled}$. In a second experiment the first 180°(^{13}C') pulse is shifted to the position $t_1/2 + T/2$, thus $^3J_{CC'}$ is active during the entire period T ($+t_1/2 + T/2 + T/2 - t_1/2 = T$), and the resulting signal $S_{^3J_{CC'}-active}$ is attenuated by a factor $\cos(2\pi J_{CC'}T)$. Then, the value of $^3J_{CC'}$ is calculated by:

$$(S_{decoupled} - S_{active})/S_{decoupled} = 1 - \cos(2\pi J_{CC'}T) = 2\sin^2(\pi J_{CC'}T) \tag{2.7}$$

2.1.4.2 ^{15}N spin echo constant time HSQC

The ^{15}N spin echo constant time HSQC experiment [189] allows the determination of the coupling constant $^3J_{CN}$ between the backbone nitrogen and the methyl carbons $^{13}CH_3$ of valine, isoleucine ($C_{\gamma 2}$), and threonine residues. The pulse sequence for the amide proton and aliphatic carbons of the experiment is identically as described above, the carbonyl spins are decoupled during the whole constant time period T by a 180°(^{13}C') pulse after $t_1/2$, and off-resonance effects caused by this pulse is refocused by another 180°(^{13}C') pulse at the end of the time period sequence. In a first experiment the $^3J_{CN}$ is refocused by a 180°(^{15}N) pulse applied at the same position as the first 180°(^{13}C') pulse (after $t_1/2$) and the signal is stored as $S_{decoupled}$. In a second experiment the 180°(^{15}N) pulse is shifted to position $t_1/2 + T/2$, causing the $^3J_{NC'}$ coupling to be active during the entire period T. The calculation of $^3J_{NC'}$ coupling is performed as before using equation 2.7.

2.1.5 NOESY experiments

2.1.5.1 3D ^{15}N and ^{13}C edited NOESY-HSQC

In heteronuclear edited experiments cross peak between ^{1}H spins are resolved according to the chemical shifts of the heteronuclei directly bounded to that protons. The ^{15}N edited NOESY-

HSQC experiment [205] allows extraction of NOE information between backbone amid H^N (but also of glutamine and asparagine side chain amide groups $-NH_2^{\delta,\epsilon}$, arginine side chain H^{N_ϵ}/N^ϵ and H^{N_ζ}/N^ζ, as well as tryptophan indole H^{N_ϵ}/N^ϵ) functions and their surrounding dipolar coupled ^{1}H partners. The pulse sequence starts with an homonuclear NOESY element, whereas during the evolution time t_1 heteronuclear coupling is refocused by an 180°(N) pulse. During the mixing time τ_m magnetization is transfered via proton homonuclear dipolar coupling. Subsequent a HSQC pulse sequence is following, in which the last 90°(H) pulse of the NOESY element correspond the first pulse of an HSQC experiment. Magnetization of protons H_n not directly bounded to the heteronucleus is suppressed by an isotope filter. In short the magnetization transfer pathway is given by:

$$\sum_{n=1}^{K} H_n(t_1) \xrightarrow{NOE} H^N \xrightarrow{^1J_{H^N N}} N(t_2) \xrightarrow{^1J_{H^N N}} H^N(t_3)$$

where K is the number of dipolar coupling partners of H^N.
Water suppression is achieved using the WATERGATE technique [139].
^{13}C edited NOESY-HSQC experiment indicate NOE connectivities between aliphatic and between aliphatic and aromatic protons. Pulse sequence is analog as described before, but composite pulse decoupling is used during t_1 and selective $^{13}C'$ decoupling is applied during t_2 (to prevent evolution of $J_{cc'}$). The transfer pathway is:

$$\sum_{n=1}^{K} H_n(t_1) \xrightarrow{NOE} H^{C_{aliphatic,aromatic}} \xrightarrow{^1J_{CC}} C_{aliphatic,aromatic}(t_2) \xrightarrow{^1J_{CC}} H^{C_{aliphatic,aromatic}}(t_3)$$

Because of better resolution in the (direct) ω_3 dimension proton chemical shift values were adopted from these spectra for the final assignment list.

2.1.6 Relaxation experiments

Backbone dynamics were characterized by measurement of backbone nitrogen ^{15}N T_1, T_2 and amide amide $^1H^N$ steady state NOE values by experiments [39] (employing sensitivity enhanced pulse sequences with pulsed field gradient for artifact suppression and coherence selection), which are described below.

2.1.6.1 T_1 Experiment

After a refocused INEPT transfer ^{15}N nitrogen transverse in-phase N_x magnetization is converted to longitudinal $-N_z$ by a 90$^\circ_y$(N) pulse. In a subsequent relaxation delay τ the inverted longitudinal magnetization recover back to its equilibrium position. During τ relaxation times amide selective 180°(^{1}H) pulses are applied [76] (which do not disturb water magnetization) in order to eliminate effects of cross correlation between ^{1}H-^{15}N dipolar and chemical shift anisotropy (CSA) relaxation mechanism. Subsequent longitudinal ^{15}N magnetization is converted into transverse magnetization by a 90°(^{15}N) pulse. Inserted into a refocused reversed INEPT transfer sequence ^{15}N frequency labeling is performed during t_1, in which a 180°(^{1}H) pulse refocuses evolution of $J_{H^N N}$ couplings. A serial of cross-peaks intensities obtained from a set of varying relaxation delays τ was fit to a two parameter decaying exponential of the form:

$$I(t) = A \exp[-(\tau/T)] \tag{2.8}$$

where $I(t)$ is peak height, and T the sought relaxation time.

2.1.6.2 T_2 Experiment

The pulse sequence of the T_2 experiment is exactly the same as for the T_1 experiment with exception of the relaxation period τ, at which a *Carr-Purcell-Meiboorn-Gill* (CPMG) spin echo pulse train is applied to determine the transverse relaxation time T_2. The CPMG sequence refocuses the contribution to relaxation value T_2 caused by inhomogeneous line broadening, and includes selective amide 180°(^1H) pulses to eliminate effects of cross correlation between ^{1}H-^{15}N dipolar and chemical shift anisotropy (CSA) relaxation mechanism. From a set of T_2 experiments the relaxation time T_2 is obtained from fitting the measured peak heights to the two parameter function given in equation 2.8 (ignoring molecular diffusion).

2.1.6.3 ^{1}H-^{15}N NOE experiment

The experiment starts with a proton $^1H^N$ saturation period of time sufficient to establish the NOE enhancement on the longitudinal ^{15}N magnetization (N_z). Subsequent frequency labeling is performed during t_2 and ^{15}N magnetization is transfered to attached amide $^1H^N$ function using a reverse INEPT sequence element and the signal is stored as S_{NOE}. In an analog manner, a reference experiment is acquired without proton $^1H^N$ pre-saturation, whose cross

peak intensities are proportional to the equilibrium magnetization of the system. The signal is stored as S_{noNOE}.

For each cross-peak of the heteronuclear steady state NOE $^{15}\mathrm{N}\{^{1}\mathrm{H}^{N}\}$ attenuation factors can be obtained from the ratio I_{NOE}/I_{nonNOE}, where I_{NOE} and I_{noNOE} is the cross peak intensity in the saturated and reference spectrum, respectively.

2.1.7 Processing

Elaborative processing of the raw NMR data is a critical step in order to extract as much information as possible from the resulting spectra.

2.1.7.1 Linear Prediction

This method is used to extend the time domain data or to remodel parts of it [6]. The linear prediction (LP) algorithms predict the k^{th} data point of the damped sinusoid NMR Signal as linear function of the previous or subsequent sampled points. Spectral distortions arising from delayed acquisition and non-linearities of the receiver can be corrected by removing the first points of the FID and applying backward prediction. Forward LP is widely used in the indirect detected dimensions of n-dimensional spectra, whose data records are almost always truncated (because of limited overall experimental time or fixed constant time delay), in order to avoid truncation artifacts or to increase the resolution.

2.1.7.2 Low frequency subtraction in the time domain

In that processing step [184] a polynomial of the n^{th} order is subtracted from the raw time signal in order to remove low frequency contribution resulting from non-suppressed water.

2.1.7.3 Zero filling

A row of zeros can be appended to a time domain data sequence prior Fourier transformation [36]. This can be done in order to fulfill the criteria necessary for applying the fast Fourier transform algorithm [35], but also to improve the information content (resolution) of a spectrum. This is because discrete Fourier transform results in a loss of the deterministic relationship of the real and imaginary components to each other (see Kramers-Koenig relation [179]), unless the raw time domain data sequence is not extended by a factor of 2. Additional zero filling

lead only to a cosmetic interpolation within the time domain data without generating further information. However, care must be taken that the FID has been brought smoothly to zero via apodization (see below) before the sequence of zero attachment starts, otherwise the resulting peak in the spectrum would appear hardly distorted.

2.1.7.4 Apodization

The process of windowing or apodization [64] is applied in order to remove truncation artifacts, increase a low signal to noise (S/N) levels and/or limited spectral resolution, or amending bad peak shapes. For that purpose the spectrum is convoluted with an appropriate line shape function $H\omega$. In practice this is facilitated by multiplying the time domain data with a filter function $h(t)$ (window function). Window function applied are sine bells or squared sine bells, whose offset and exact shape is chosen to obtain best results. More weighting the first part of the FID results in an increased S/N ratio by (unfortunately) raising the line width, whereas weighting of the later part give rise for an improved resolution. An essential feature of the implementation of a window function is the smoothly attenuation of the FID to zero prior end to the time domain t_{end}. Otherwise the transformed spectrum $S_{freq}(\omega)$ of the truncated FID would be convoluted with a function $sinc(t_{end}\omega)$, resulting in peak shapes that offer disadvantageous sinc wiggles (oscillating truncation artifacts).

2.1.7.5 Fourier transformation

The conversion of the digitally sampled time domain data (FID) into the frequency domain (spectrum) can be obtained by discrete *Fourier transformation*:

$$S_{freq} = S(l/N\Delta t) = \mathfrak{F}S_{FID}(k\Delta t) = \sum_{k=0}^{N-1} S_{FID}(k\Delta t)e^{-i2\pi kl/N} \tag{2.9}$$

where N is the number of complex data points, Δt the sampling interval, and $l = -N/2, ..., 0, ..., +N/2$. In practice the discrete Fourier transform is rather performed using the fast Fourier transformation (FFT) algorithm [35], than executing the direct numerical calculation. However, for this method, the number of data points N have to be an integral power of 2 (i.e. $N = 2^n$, n is an integer). Otherwise the data sequence must be extended to a power of 2 either per zero filling or linear prediction.

2.1.7.6 Phasing

For best performance NMR data interpretation, the peaks of the spectra should offer pure double absorption lineshapes. Dispersion modes give rise for a skewed appearance of the a spectrum with twisted lineshapes and would dramatically decrease the spectral resolution. Phase errors in the direct dimension are introduced by spectrometer imperfections, whereas phase shifts of the indirect dimensions arising from the pulse sequence. For visualization, the cosine and sine modulated (both acquired for quadrature detection in the indirect ω_1 dimension) time domain signal related to a spin which has been evolved under different chemical shift Hamiltonians during evolution time t_1 and acquisition time t_2 can be written as:

$$s_c(t_1, t_2) = \cos(\Omega_1 t_1) e^{i\Omega_2 t_2} e^{i\phi_2} \tag{2.10}$$

$$s_s(t_1, t_2) = \sin(\Omega_1 t_1) e^{i\Omega_2 t_2} e^{i\phi_2} \tag{2.11}$$

Fourier transform in ω_2 dimension gives:

$$\begin{aligned} s_c(t_1, \omega_2) &= \cos(\phi_2)\cos(\Omega_1 t_1) A_2 - \sin(\phi_2)\cos(\Omega_1 t_1) D_2 \\ &+ i(\sin(\phi_2)\cos(\Omega_1 t_1) A_2 + \cos(\phi_2)\cos(\Omega_1 t_1) D_2) \end{aligned} \tag{2.12}$$

$$\begin{aligned} s_s(t_1, \omega_2) &= \cos(\phi_2)\sin(\Omega_1 t_1) A_2 - \sin(\phi_2)\sin(\Omega_1 t_1) D_2 \\ &+ i(\sin(\phi_2)\sin(\Omega_1 t_1) A_2 + \cos(\phi_2)\sin(\Omega_1 t_1) D_2) \end{aligned} \tag{2.13}$$

where A_x and B_x represent absorption and dispersion Lorentzian's at frequency x in F_2 dimension, respectively. Phase adjustment in is facilitated by generating linear combination of these two data sets:

$$s_{real}(t_1, \omega_2) = \mathsf{Re}[s_c(t_1, \omega_2)] + i\mathsf{Re}[s_s(t_1, \omega_2)] \tag{2.14}$$

$$s_{imaginary}(t_1, \omega_2) = \mathsf{Im}[s_c(t_1, \omega_2)] + i\mathsf{Im}[s_s(t_1, \omega_2)] \tag{2.15}$$

Fourier transform in indirect ω_1 dimension gives:

$$s_{real}(\omega_1, \omega_2) = \cos\phi_2 A_1^+ A_2 - \sin\phi_2 A_1^+ D_2 + i(\cos\phi_2 D_1^+ A_2 - \sin\phi_2 D_1^+ D_2) \tag{2.16}$$

$$s_{imaginary}(\omega_1, \omega_2) = \sin\phi_2 A_1^+ A_2 - \cos\phi_2 A_1^+ D_2 + i(\sin\phi_2 D_1^+ A_2 - \cos\phi_2 D_1^+ D_2) \tag{2.17}$$

where A_x^s and B_x^s represent absorption and dispersion Lorentzian's at frequency x and position

s in F_1 dimension, respectively.

Pure double absorption line shape are obtained by a proper $\cos\phi_{2,emp}$ and $\sin\phi_{2,emp}$ combination of the real parts of $s_{real}(\omega_1,\omega_2)$ and $s_{imaginary}(\omega_1,\omega_2)$ (equation 2.17), respectively:

$$\begin{aligned}\cos\phi_{2,emp}\mathsf{Re}[s_{real}(\omega_1,\omega_2)] + \sin\phi_{2,emp}\mathsf{Re}[s_{imaginary}(\omega_1,\omega_2)] = \\ \cos\phi_{2,emp}[\cos\phi_2 A_1^+ A_2 - \sin\phi_2 A_1^+ D_2] + \sin\phi_{2,emp}[\sin\phi_2 A_1^+ A_2 - \cos\phi_2 A_1^+ D_2] = \\ A_1^+ A_2(\mathsf{if}\phi_{2,emp} \approx \phi_2) \qquad (2.18)\end{aligned}$$

whereas $\phi_{2,emp}$ is empirically adjusted to minimize the phase error in the spectra ($\phi_{2,emp} \approx \phi_2$). The term $\phi_{2,emp}$ can contain a frequency independent zero order phase correction $\phi_{2,emp}^{zero}$ and a frequency dependent first order correction $\phi_{2,emp}^{st}$ contribution in order to achieve best results for the individual resonances in the spectrum.

If there is also a phase error in t_1 present, the time domain signals change to:

$$s_c(t_1,t_2) = \cos(\Omega_1 t_1 + \phi_1)e^{i\Omega_2 t_2}e^{i\phi_2} \qquad (2.19)$$

$$s_s(t_1,t_2) = \sin(\Omega_1 t_1 + \phi_1)e^{i\Omega_2 t_2}e^{i\phi_2} \qquad (2.20)$$

Assuming that ϕ_2 is already corrected, after real (cosine) Fourier transform in ω_2 dimension one get for $s_{real}(t_1,\omega_2)$:

$$s_c(t_1,\omega_2) = e^{i\Omega_1 t_1 + \phi_1}A_2 \qquad (2.21)$$

Fourier transform in ω_1 dimension results in:

$$s_c(\omega_1,\omega_2) = e^{i\Omega_1}(A_1^+ + iD_1^+)A_2 \qquad (2.22)$$

The pure lineshapes $A_1^+ A_2$ are generated by applying a phase correction of $e^{-i\Omega_{1,emp}}$ in the ω_1 dimension.

2.1.7.7 Baseline correction

For peak recognition and correct integration a flat base plane in the frequency domain is obligate. For that purpose several algorithm [64] exists, which dry to identify and fit the baseline in rows of spectra containing no signal before global baseline correction is performed.

2.1.7.8 Referencing

To facilitate comparison of resonance positions obtained from different samples, spectra and spectrometers, the measured chemical shifts are referenced against a standard compound. In protein NMR, 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS, $C_6H_{16}O_3SSi$, M_w=196.34, δ^1H = 0.0 ppm) is commonly used as an internal chemical shift standard for protons. Because of the insensitivity of the heternuclei (^{13}C,^{15}N), an indirect calibration of their chemical shifts is applied using the relationship:

$$\nu_0^X = \nu_0^H \gamma_X / \gamma_H \tag{2.23}$$

where ν_0^H and ν_0^X is the absolute frequency of 0 ppm for the proton and the X nucleus, respectively, and γ_X/γ_H is their gyromagnetic ratio. In this work, values of 0.25144952 and 0.10132905 were used for the ratios $\gamma_{^{13}C}/\gamma_H$ and $\gamma_{^{15}N}/\gamma_{^1H}$, respectively. By doing so, the experimentally required heteronuclear offset [MHz] is calculated as $\nu_0^X(1 + \cdot c10^{-6})$, where c is the related ppm value.

2.1.8 Resonance assignment strategy

2.1.8.1 Sequential backbone chemical shift assignment

In conventional (indirect) NMR conformational analysis each nuclear magnetic resonance have to be related to a specific spin in a particular amino acid of the protein under investigation prior NOE assignment and derivation of distance restraints for subsequent structure calculation can take place.

For that purpose, pairs of triple resonance experiments using relatively uniform and well resolved one and two bond heteronuclear couplings are exploited for sequential correlation of $^1H^N$, ^{15}N, $^{13}C^\alpha$, and $^{13}C^\beta$ resonances. For example, the combined use of 3D HNCA, which provides intra-residue together with some sequential correlation between $^1H^N$, ^{15}N, and $^{13}C^\alpha$ resonances, and the 3D HN(CO)CA experiment, which indicate solely the inter residue correlation between the $^1H^N$ and ^{15}N resonances of a residue i and the $^{13}C^\alpha$ chemical shifts of the residue $i-1$ preceding in primary sequence, facilitates an obvious route for sequential assignment. Typically the chemical shift dispersion of the $^{13}C^\alpha$ chemical shifts is to low to generate unambiguous spinsystem fragments of sufficient length for sequential mapping. Therefore, at least two pair combinations of standard set of triple resonance experiments have normally to be acquired for successful sequence specific assignment (see figure2.1). For instance, a com-

bination of 3D HNCA/3D HN(CO)CA and 3D HNCO/3D HN(CA)CO spectra enables beside correlation of $^{1}H_{i}^{N}$, $^{15}N_{i}$, $^{13}C_{i,i-1}^{\alpha}$ chemical shifts, the use of the $^{13}C_{i,i-1}^{'}$ resonances for resolving ambiguities caused by shift degeneracy. Alternatively, the provide additional use of 3D CBCA(CO)NH (where coherence transfer at the initial INEPT step is optimized especially for $^{13}C^{\beta}$ spins to ensure complete set of $^{13}C^{\beta}$ chemical shifts) the more characteristic $^{13}C_{i,i-1}^{\alpha}$ and $^{13}C_{i,i-1}^{\beta}$ chemical shift pair for sequential linking of the backbone $^{1}H_{i}^{N}$ and $^{15}N_{i}$ chemical shifts. Because of the pronounced sensitivity of the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts on the secondary structure, this procedure is almost guaranteed to work in the chemical backbone assignment of folded proteins (provided sufficient protein solubility and stability). The situation is different in the assignment process of partly structured or flexible intrinsic unfolded proteins, where the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shift dispersion is very poor. However, the study of such proteins is likewise of particular interest when studying protein folding mechanisms, protein missfolding and aggregation, which is matter of several human diseases. Here, instead of the conformational sensitive $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts the more sequence dependent backbone $^{1}H^{N}$, ^{15}N and $^{13}C'$ chemical shifts are employed for sequential linking. The chemical shift dispersion of these resonances (particular ^{15}N and $^{13}C'$) is still good even in unfolded molecular systems. The more recently published 3D HNN [136], which correlate the $^{1}H_{i}^{N}$ and $^{15}N_{i}$ resonances of residue i with the $^{1}H_{i+1,i-1}^{N}$ and $^{15}N_{i+1,i-1}$ chemical shifts adjacent in the primary sequence, and 3D HN(C)N experiment, which solely provide the inter-residue connection of $^{1}H_{i}^{N}$ and $^{15}N_{i}$ spins to the subsequent $^{1}H_{i+1}^{N}$ and $^{15}N_{i+1}$ resonances, are very useful in such cases. Their spectral characteristics enable direct identification of certain triplets of residues (ZGY, GXY, GGY, and ZXY; where G is glycine and X,Y,Z is any other residue type than glycine or proline). For example, diagonal peaks of gylcines and cross peaks to that residue type from others exhibit opposite signs (see figure 2.2). Prolines indicate their presence in an absence of a peak. Because in pulse sequence the magnetization remains on the $^{13}C^{\alpha}$ spin for a long time, the inherent substantial loss of signal is only tolerable in samples of flexible unfolded proteins (where transverse $^{13}C^{\alpha}$ relaxation times are long enough for reasonable results).

In cases of folded proteins sequential $^{1}H^{N}$-$^{1}H^{N}$, $^{1}H^{N}$-$^{1}H^{\alpha}$ and $^{1}H^{N}$-$^{1}H^{\alpha}$ NOEs from the 3D ^{15}N-edited NOESY-HSQC spectrum can be additionally exploited in this work to decrease the ambiguity of the assignment process.

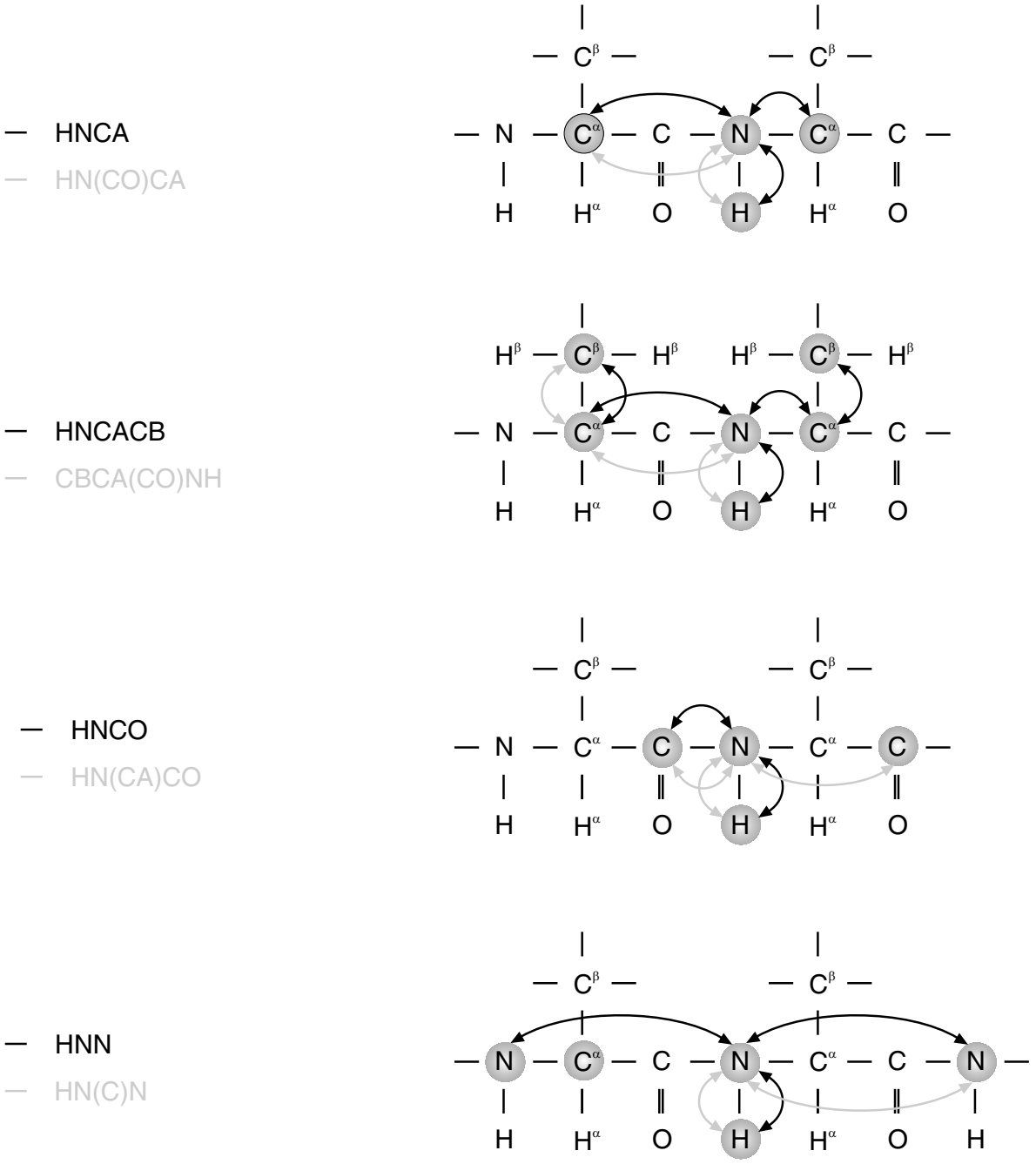


Figure 2.1: Suite of standard triple resonance experiments employed for sequential backbone chemical shift assignment. Provided correlations between various backbone spins (labeled with circles) are indicated by arrows.

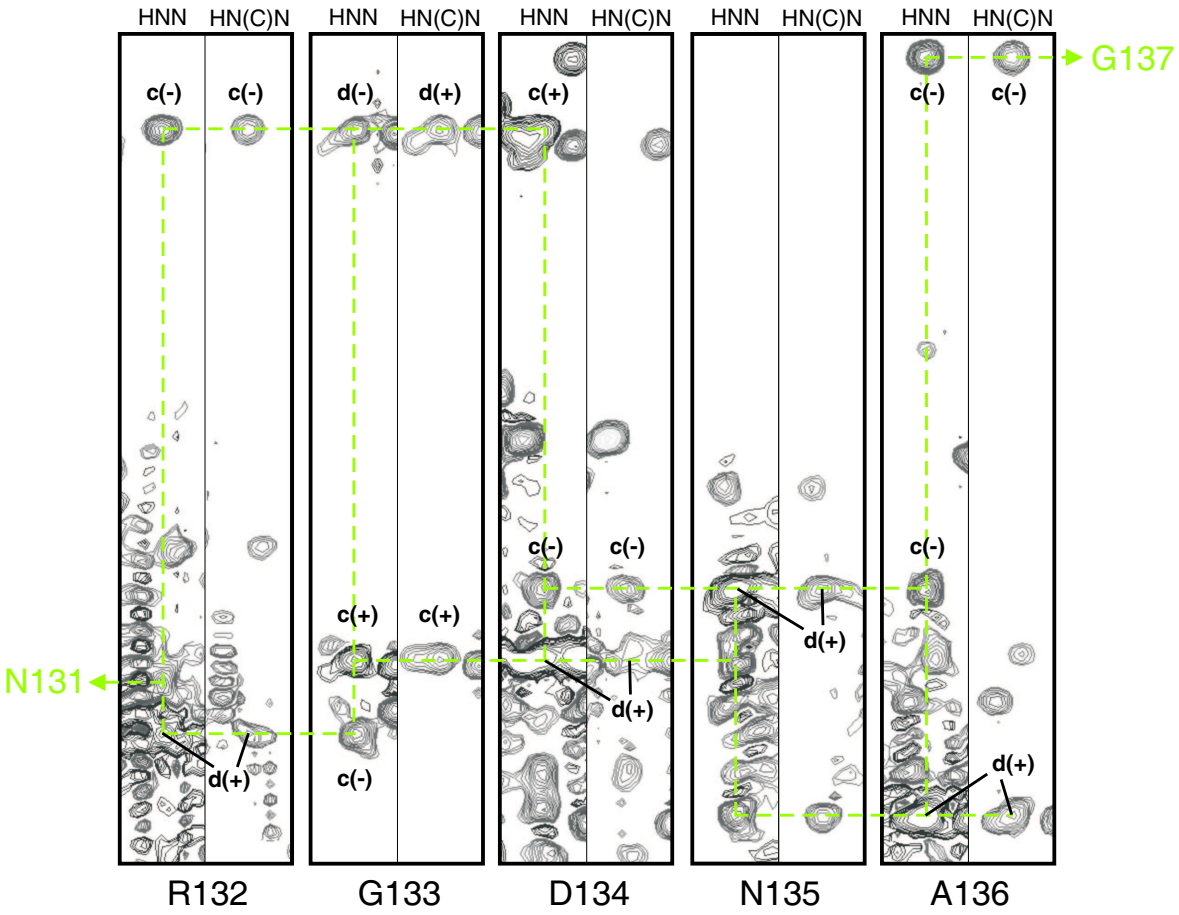


Figure 2.2: Observed peak patterns for a stretch of fi ve residues of brain abundant protein 1 (BASP1). Diagonal and cross peaks are labeled with **d** and **c**, respectively, whereas their relative intensity signs are indicated within the brackets $(+,-)$.

2.1.8.2 Identification of secondary structure elements

Different secondary structure elements give rise for characteristic NOE patterns [126]. In α-helices (especially featured by NOEs between residues i and $i-3$, and between residue i and $i-4$, respectively) the following set of NOE contacts are potentially found (intensity attributes are given in the brackets):

- sequential $^{1}\mathrm{H}_{i}^{N}$-$^{1}\mathrm{H}_{i-1}^{N}$ (strong)
- medium range $^{1}\mathrm{H}_{i}^{N}$-$^{1}\mathrm{H}_{i-2}^{N}$ (medium)
- sequential $^{1}\mathrm{H}_{i}^{N}$-$^{1}\mathrm{H}_{i-1}^{\alpha,\beta}$ (medium-strong)
- medium range $^{1}\mathrm{H}_{i}^{N}$-$^{1}\mathrm{H}_{i-3}^{\alpha}$ (medium)
- medium range $^{1}\mathrm{H}_{i}^{N}$-$^{1}\mathrm{H}_{i-4}^{\alpha}$ (weak)

Specific NOEs commonly observed in β-sheets are:

- sequential $^{1}\mathrm{H}_{i}^{N}$-$^{1}\mathrm{H}_{i-1}^{\alpha}$ (strong)
- sequential $^{1}\mathrm{H}_{i}^{N}$-$^{1}\mathrm{H}_{i-1}^{\beta}$ (medium)
- inter-strand long-range $^{1}\mathrm{H}_{i}^{N}$-$^{1}\mathrm{H}_{j}^{N}$ (medium)
- inter-strand long-range $^{1}\mathrm{H}_{i}^{N}$-$^{1}\mathrm{H}_{j}^{\alpha}$ (medium)
- inter-strand long-range $^{1}\mathrm{H}_{i}^{\alpha}$-$^{1}\mathrm{H}_{j}^{\alpha}$ (strong in anti-parallel β-sheets, medium-weak in parallel β-sheets)

Beside characteristic NOE patterns, secondary elements can be identified from secondary chemical shifts:

$$\Delta\delta X = \delta X_{experimental} - \delta X_{randomcoil} \tag{2.24}$$

where X denotes the commonly employed nuclei used for calculation ($^{1}\mathrm{H}^{\alpha}$, $^{13}\mathrm{C}^{\alpha}$, $^{13}\mathrm{C}^{\beta}$, and ^{13}C'). Secondary chemical shifts exhibits different signs depending on the spin and the secondary structure element. For α-helices the $\Delta\delta X$ value is positive for $^{13}\mathrm{C}^{\alpha}$ spins and negative for $^{13}\mathrm{C}^{\beta}$ spins. In β-sheets, the reverse being present. Therefore, the expression $^{13}\mathrm{C}^{\alpha}-^{13}\mathrm{C}^{\beta}$ provides a sensitive measure for detecting the presence or preference of secondary structure.

At mean, the $\Delta\delta X$ values are found to be more distinct for α-helices than for β-sheets. In order to calculate secondary chemical shifts depending on the conformation specifying backbone ϕ,ψ torsion angles only, the applied random coil chemical shift should be corrected for local sequence effects [163]. This is particularly important for evaluation of secondary element preferences in partly folded or unfolded proteins, where the observed secondary chemical shifts are small.

2.1.8.3 Side chain assignment

Correlation from the backbone $^1H^N$ and ^{15}N chemical shifts to the side chain aliphatic resonances is facilitated by ^{1}H,^{13}C relayed TOCSY experiments. In this work the 3D H(CCCO)NH and 3D (H)CC(CO)NH experiments were applied to achieve this goal (see figure 2.3. However, due to relaxation not all correlations can be observed in practice, especially for leucine, isoleucine, argenine and lysine sidechain resonances, for which magnetization transfer is rather time-consuming. However, additional (if not most) information is inferred from the 3D ^{13}C-edited NOESY-HSQC, at which the proton sidechain chemical shifts are extracted more accurately from the better resolved direct ω_3 dimension. Predominant intra-residue NOEs can be observed for sidechain methylene protons, which alleviate the resonance assignment of longer sidechains. Most intense NOE correlation of $^1H^\delta$ protons in the 3D ^{13}C-edited NOESY-HSQC spectrum are generally from the intra-residue $^1H^\beta$ protons because of close spatial proximity ($\leq$ 2.8Å). These NOEs facilitates the linking of the aromatic sidechain protons to the (non-scalar coupled) aliphatic $^1H^\beta$ resonances of residues phenylalanine and tyrosine, but also the connectivity of heterocyclic aromatic protons to the $^1H^\beta$ chemical shifts of residues tryptophan and histidine. Sidechain NOE contacts for the δ-NH_2 and ϵ-NH_2 groups of residue asparagine and residue glutamine to their sidechain methylene -CH_2^β and -CH_2^γ functions, respectively, as well as for the indolic $NH^{\epsilon 1}$ proton can be observed in the 3D ^{15}N-edited NOESY-HSQC.

2.1.9 Ligand binding in NMR

In a simple case, the interaction of a ligand L to a protein P can be formulated as a two site second order exchange:

$$L + P \rightleftharpoons LP \tag{2.25}$$

Beside the location of the ligand binding site on a protein surface knowledge about the stability of the complex is fundamental. The affinity of a ligand to a protein is quantified by the temper-

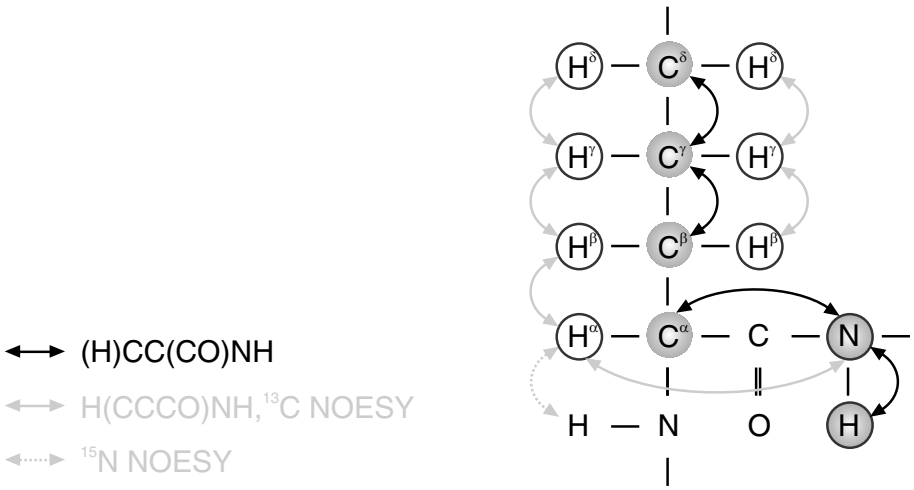


Figure 2.3: Through bond (black arrows) and through space (gray arrows) correlation experiments applied for sequential sidechain chemical shift assignment.

ature dependent equilibrium dissociation constant of the protein-ligand complex K_D, which is given by the ratio between the k_{off} and the k_{on} rate:

$$K_D = \frac{k_{off}}{k_{on}} = \frac{[L] \cdot [P]}{[LP]} \tag{2.26}$$

where $[P]$ denotes the molar concentration of free (i.e. unbounded) protein, $[L]$ the concentration of the free ligand, and $[LP]$ the concentration of the protein ligand complex. The unimolecular rate constant k_{off} is inversely proportional to the mean lifetime τ_{PL} of the protein-ligand complex. The bimolecular rate constant k_{on} is a measure of the probability of a productive encounter between free protein and ligand. The total concentration of the protein and the ligand are denoted by $[P_0]$ and $[L_0]$, respectively:

$$[P_0] = [P] + [PL] \qquad [L_0] = [L] + [PL] \tag{2.27}$$

Combining equation 2.26 and 2.27, the concentration of the protein ligand complex can be expressed in terms of the measurable total protein and ligand concentration and the equilibrium constant:

$$[PL] = 1/2\{([P_0] + [L_0] + K_D) - (([P_0] + [L_0] + K_D)^2 - 4 \cdot [P_0][L_0])^{1/2}\} \tag{2.28}$$

If the lifetime τ_{PL} of the complex is short compared with the chemical shift differences $\Delta\delta$ between the free and ligand bounded state ($K_D \gg \Delta\delta$), then fast exchange occurs and a single

weighted averaged resonance is observed for the protein nuclei involved in ligand binding:

$$\delta_{obs} = \delta_P \cdot p_P + \delta_{PL} \cdot p_{PL} \tag{2.29}$$

where δ_{obs} is the observed chemical shift, δ_P and δ_{PL} are the chemical shifts of the free protein and of the protein in the ligand bounded state, respectively, and p_P and p_{PL} are the corresponding mole fractions. In the course of a titration, the mole fractions p_P and p_{PL} change, and therefore the observed chemical shift of the protein nuclei involved in ligand binding in a hyperbolic manner (figure). The chemical shift difference $\Delta\delta$ is proportional to the mole fractions of the protein in the ligand bounded state:

$$\Delta\delta = |\delta_{obs} - \delta_P| = \delta_P \cdot (p_P - 1) + \delta_P p_{PL} = p_{PL}(\delta_{PL} - \delta_P) \tag{2.30}$$

Using $p_{PL} = [PL]/[P_0]$, the chemical shift change upon ligand addition is given by:

$$\Delta\delta = \frac{1}{2}\{([P_0] + [L_0] + K_D) - (([P_0] + [L_0] + K_D)^2 - 4 \cdot [P_0][L_0])^{1/2})\}\frac{\delta_{PL} - \delta_P}{[P_0]} \tag{2.31}$$

Thus, the values of K_D and can be determined from a titration experiment by non linear least square fitting of the chemical shift changes $\Delta\delta$ as a function of ligand concentration $[L_0]$. This can be performed for all residues for which chemical shift changes are detected.

2.1.10 Relaxation, molecular motion and spectral density mapping

NMR relaxation is ideally suited to sample, on a residue to residue basis, the whole range of molecular motions spanning from picoseconds (typically molecular vibrations and rotations) to seconds (also containing chemical exchange and molecular diffusion). Classically, the dynamic behavior is assessed from measurements of heteronuclear longitudinal T_1 and transverse T_2 relaxations times as well as the steady state ^{1}H-^{15}N NOE, which specify the motional properties of the backbone N-H vectors. Studies on native proteins have suggested that the majority of the main chain sites are highly restricted to small amplitude motions on the picosecond to nanosecond time scale, while the mobility of the residues located in loops and near polypeptide chain termini of proteins can be more substantial. Hydrophobic packing, hydrogen bonding and residual burial were indicated as the principle determinants of molecular mobility. While the native states of proteins were dynamical characterized extensively, relatively less data exists for partially folded or intrinsic unfolded proteins. However, the remarkable percentage of the

proteins in vital cells are found to be generally inherent unfolded. Although some of them need to be in complex with other factors or ligands (which e.g. serves as neutralizer of repulsive forces preventing folding) to get folded, intrinsic disordered proteins can exhibit inherent preferences for secondary structure elements. Whilst backbone chemical shifts, coupling constants and NOE intensities provide an indication of the existence of more ordered regions and their relative populations, only relaxation studies allow a description of the motional time scale and the extent of the motions involved.

The relaxation mechanism for the N-H spinsystem arises from the local time-varying magnetic fields generated at the ^{1}H and ^{15}N nuclei due to global tumbling and internal mobility of the various N-H vectors in the molecule. For these spin $1/2$ nuclei, the predominant relaxation mechanism are dipole-dipole interaction and usually to an lower extent the chemical shift anisotropy (CSA). The frequency spectrum of the rotational motions of the N-H vector in a molecule relative to the external superconducting magnetic field is given by the spectral density function, $J(\omega)$, which describes how the energy of the dipole-dipole and CSA interaction is distributed at the five different angular frequencies 0, ωH, $\omega N + \omega H$, ωN, and $\omega H - \omega N$. The three NMR relaxation measurements T_1, T_2 and ^{1}H-^{15}N NOE are related to $J(\omega)$ as a function these values as follows:

$$1/T_1 = R_1 = (d^2/4)[J(\omega_H - \omega_N) + 3J_{\omega_N} + 6J(\omega_H + \omega_N)] + cJ(\omega_N) \tag{2.32}$$

$$1/T_2 = R_2 = (d^2/8)[4J(0) + J_(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)] + (c^2/6)[3J(\omega_N) + 4J(0)] + R_{ex} \tag{2.33}$$

$$NOE = 1 + (d^2/4R_1)(\gamma_N/\gamma_H)[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)] \tag{2.34}$$

where $d = (\mu_0 h/8\pi^2)\gamma_H\gamma_N\langle r_{NH}^{-3}\rangle$, $c = \omega_N(\sigma_\parallel - \sigma_\perp)/\sqrt{3}$, ω_N and ω_H the lamor frequency of the nitrogen and proton nuclei, γ_N and γ_H their related gyromagnetic ratios, h is the Planck's constant, μ_0 is the permeability of the free space, r_{NH} is the nitrogen bond length chosen as 1.02 Å in this work, and $\sigma_\parallel - \sigma_\perp$ is the chemical shift anisotropy of ^{15}N (-160ppm).

Because of the conservation of the total mean dipole-dipole and CSA energy available to an ensemble of independent molecules, the area under Jw is a constant. As a consequence, the density of motions appear shifted to lower frequencies for slower tumbling structural parts or molecules. The reverse holds for higher dynamical regions. In other words, the magnitude of the spectral density function sampled at the five frequencies listed above indicates the extent of the N-H bond motions. For restricted N-H bond motions the most significant contribution will

arise from the low frequency component $J(0)$, whereas in the opposite case of high mobility the high-frequency component $J(\omega_H)$ is expected to be the main contributor.

The three measured parameters T_1, T_2 and ^{1}H-^{15}N NOE constitute only three equations for the five unknowns $J(0)$, $J(\omega_H)$, $J(\omega_N + \omega_H)$, $J(\omega_N)$, and $J(\omega_H - \omega_N)$, hence they are intrinsically incapable for determining the spectral density at any of that probing frequencies. However, in case of ^{15}N relaxation (note that $\omega_H \sim 10 \cdot \omega_N$) the spectral density function variate rather slowly around the proton lamor frequency ω_H, such that one equate the three values of the density function at frequencies ω_H, $\omega_N + \omega_H$ and $\omega_H - \omega_N$ into an averaged value $\langle J_{\omega_H} \rangle$. Furthermore, assuming that the contribution from slow chemical exchange R_{ex} is much smaller than R_2, the different relaxation parameters can be used to determine the values of spectral density functions $J(0)$, $J(\omega_H)$, and $J(\omega_N)$ as follows by:

$$J(\omega_H) = 0.2R_1(NOE - 1)(4/d^2)(\gamma_N/\gamma_H) \quad (2.35)$$

$$J(\omega_N) = 4R_1/(3d^2 + 4c^2)[1 - (7/5)(NOE - 1)(\gamma_N/\gamma_H) \quad (2.36)$$

$$J(0) = 1/(3d^2 + 4c^2)[6.R_2 - 3R_1 - (18R_1/5)(NOE - 1)(\gamma_N/\gamma_H)] \quad (2.37)$$

2.2 NMR and empirical derived structural restraints

Chemical shifts, scalar coupling interactions, and dipolar relaxation are sensitive to molecular conformation, and quantification of these parameters permits deduction of structural restraints for conformational analysis of a biomolecule by NMR spectroscopy. A set of structural restraints can comprise distance restraints (NOE, hydrogen bonds), dihedral restraints (from J-couplings or derived from cross-correlation experiments), orientational restraints (residual dipolar couplings) and others (database restraints).

2.2.1 NOE distance restraints

Up to now the most important observable parameter used in structural analysis of a protein is the NOE [126]. The initial rate of NOE enhancement buildup between two dipolar coupled spins I and S in a transient NOE experiment (such as NOESY) can be approximated as linear:

$$\frac{dI_{cross}(\tau_m)}{d\tau_m}|_{\tau_m} = -\sigma_{IS} I^0_{diagonal} \quad (2.38)$$

where dI_{cross} and $I_{diagonal}$ is the intensity (or volume) of the cross peak at time τ_m and of the

diagonal peak at time $\tau_m = 0$, respectively, and σ_{IS} the cross relaxation rate, which is inverse proportional to the inverse sixth power of the distance r_{IS} between the spins I and S. If one inter-proton distance r_{ref} is known (e.g. from a diastereotope methylene proton pair) then an unknown inter-proton distance r_x is determined by the relationship (ignoring differential internal mobility and spin diffusion):

$$r_x = r_{ref}(\frac{I_{ref}}{I_x}) \tag{2.39}$$

in which I_{ref} and I_x denotes the cross peak intensities (or volumes).

Due to r^{-6} distance dependency of the rate σ_{IS}, cross relaxation proceeds faster at smaller values of σ_{IS}. Thus, as a result of internal mobility the calculated distances r_x are smaller than the time average $< r_x >$. In addition, the mixing time τ_m parameter of the NOESY experiment is set to a value (of the order $1/R_1$ in which R_1 is the smallest longitudinal relaxation rate in the protein) suitable to allow a significant NOE to develop. For longer mixing times τ_m, multiple magnetization transfers (spin diffusion) in a network of dipolar coupled spins will occur that substantially contributes to the cross peak intensities. Therefore, these measures are not longer directly proportional to the cross relaxation constant between the NOE related nuclei I and S. For simplification, the NOESY cross peaks are grouped according of their intensities into several distance ranges (e.g. strong, medium, weak, and very weak).

The upper bound distance is estimated from NOE intensities observed for protons of known covalent geometry or inter-proton distances commonly found in regions of regular secondary structure (e.g. sequential H_α-H^N distances in β strands or H^N-H^N distances in α-helices). The lower bound values are typically set according the sum of the van der Waal radii of the distance related atoms.

Generally, NOE distance restraint generation is complicated by missing peaks (incomplete picking or signal inexistent because of fast relaxation), peak pick errors in crowded spectral regions, artifacts of spectral processing, incomplete water suppression, sample impurities, noise, splitting of peaks caused by strong coupling, or the presence of multiple protein conformations. It takes several rounds of NOE assignment and subsequent structure calculation to obtain a self-consistent network of unambiguous distance restraints.

The derived NOEs can be classified according the distance of their related atoms i and j in the primary sequence:

- intra-residue NOEs
- sequential inter-residue NOEs ($|i - j| = 1$)

- medium-range inter-residue NOEs ($2 \leq |i-j| \leq 4$)
- long-range inter-residue NOEs ($|i-j| > 4$)

Of particular interest are long-range NOE contacts, since they define the global fold of the protein. For pro-chiral methyl carbons, pro-chiral methylene protons (α-CH_2 protons in glycines, β-, γ-, and δ-CH_2 protons of others) with missing stereo specific assignment, for methyl protons and protons of aromatic rings (that rotate fast on the NMR chemical shift time scale) a centrally located pseudo atom is used for the distance restraint generation.

2.2.2 Dihedral angle restraints

Chemical shifts of backbone atoms and three bond couplings in proteins contain valuable information about the local conformation. Especially, the secondary chemical shifts of $^1H^\alpha$, $^{13}C^\alpha$, and $^{13}C^\beta$ were found to correlate closely with its Φ and Ψ torsion angles. A feature of backbone amid H^N chemical shifts are there periodicity observed in many α-helical structures, and their distinct dependence on hydrogen bond length, found in contrast to backbone C' chemical shifts values. From intra-residue three bond ^{13}C-^{13}C J and ^{15}N-^{13}C J couplings structural information for sidechain χ_1 torsion angles of valine, isoleucine and threonine, and sidechain χ_2 torsion angles of leucine could be extracted, which in turn also facilitates the stereo specific assignment of the methyl groups of leucine and valines.

2.2.2.1 Prediction of backbone Φ and Ψ torsions from secondary chemical shift

Homologous proteins display quite similar patterns of backbone secondary chemical shifts [148]. The program TALOS [25] uses the inverse of this relation to predict the most likely backbone Φ,Ψ torsion angles by search a database for adjacent residues with secondary chemical shift and sequence similarity. The program is written in the tcl/tk language, take use of NMRWish, a companion package of the NMRPipe analysis system [29], and provides a nice graphical user interface for examining the results. Its database contains a nearly complete chemical shift assignment of now about 200 of proteins, for which crystal structures are available solved at low resolution ($\leq$ 2.2Å). The program reads as input the backbone chemical shifts and the primary sequence and evaluates the experimental secondary chemical shifts and the sequence similarity for a string of three sequential amino acids i of the query protein relative to all triplets of sequential amino acids j contained in the database. A similarity factor $S(i,j)$ is calculated

by:

$$
\begin{aligned}
S(i,j) &= \sum_{n=-1}^{+1} [k_n^0 \Delta^2_{ResType} \\
&+ k_n^1(\Delta\delta C^{\alpha}_{i+n} - \Delta\delta C^{\alpha}_{j+n})^2 \\
&+ k_n^2(\Delta\delta N_{i+n} - \Delta\delta N_{j+n})^2 \\
&+ k_n^3(\Delta\delta C^{\beta}_{i+n} - \Delta\delta C^{\beta}_{j+n})^2 \\
&+ k_n^4(\Delta\delta C'_{i+n} - \Delta\delta C'_{j+n})^2 \\
&+ k_n^5(\Delta\delta H^{\alpha}_{i+n} - \Delta\delta H^{\alpha}_{j+n})^2] \qquad (2.40)
\end{aligned}
$$

$\Delta\delta$ are the the secondary chemical shifts for the $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$, $^{1}H^{\alpha}$, and ^{15}N nuclei. The factor $\Delta_{ResType}$ is an integer number describing how similar two compared residue types are (ranging from 0 (high similarity) to 3 (low similarity)). Examples for high identities are residues with aromatic sidechains, or C^{β}-branched residues (isoleucine, leucine, ...) typically found in β sheets. In contrast, residues with 'solo' characters are glycines (frequently offering positive Φ angles) and prolines showing a very restricted range of Ψ,Φ torsions. The individual weighting factors k_n^x (x = 1-5) are statistically derived from the database and are introduced to provide a measure how meaningful a given type of secondary chemical shift (e.g. $\Delta\delta N_{i-1}$) is for predicting the ψ,ϕ angles of a residue i. For obtaining best results (minimizing number of erroneous predictions) these factors were subsequently refined by empirical adjustment.

The program then reports the 10 best matches of all database triplets j for a residue i with lowest $S(i,j)$ values, and classifies those predictions as 'unambiguous', where at least 9 ψ,ϕ angles of the 10 best rapports fall into the same populated region of the Ramachandran plot, and none of the center residues in the 10 strings offers a positive Φ angle. Unambiguous accepting of positive Φ angle predictions are not performed because of small occurrence in the database. Afterwards, the user can check and define further predictions as 'unambiguous' by additional use of NOESY backbone H^N/H^{α} patterns and intensity ratios, or $^3J_{H^N H^{\alpha}}$ couplings, if available. For comparison, the ψ,ϕ torsions of a preliminary structure or a homology model can be viewed in the Ramachandran map to make the decisions of accepting a prediction easier. In the course of subsequent structure calculation and examination of the TALOS results, further predictions can be classified as 'unambiguous' and converted to torsion restraints.

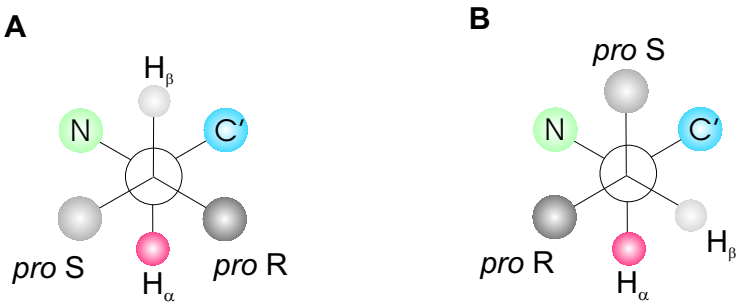


Figure 2.4: Derivation of (A) *trans* or (B) *gauche* confi guration along sidechain χ_1 axis by dint of an experimental $^3J_{C_\gamma N}/^3J_{C_\gamma C'}$ coupling pair. See text for explanation.

2.2.2.2 Deriving χ_1 torsions restraints for valine, isoleucine, and threonine residues from experimental $^3J_{C_\gamma N}/^3J_{C_\gamma C'}$ couplings

As the magnitude of the $^3J_{C_\gamma N}/^3J_{C_\gamma C'}$ couplings is a function of the dihedral angle χ_1, their measurement provides conformational information. In addition, the knowledge of $^3J_{C_\gamma N}$ and $^3J_{C_\gamma C'}$ facilitates the stereospecific assignment of the methyl groups of valines and leucines. Generally, the translation between coupling constants and the dihedral angle Θ is accomplished through an empirical Karplus relation usually given by:

$$^3J(\Theta) = A\cos^2\Theta + B\cos\Theta + C \tag{2.41}$$

The coefficients A, B, and C are sensitive on both the nature of the coupled nuclei and the local chemical environment. In contrast to backbone ψ,ϕ torsions, for analysis of the side-chain torsion angle χ_1 only a few Karplus curve parameterizations for heteronuclear couplings are available [143]. Because of this, restraints for χ_1 torsions are derived in a more semi-quantitative manner. ^3J couplings of larger size value (3-4 Hz) indicate *trans* configuration whereas small values ($\sim$ 1 Hz) correspond to *gauche* configurations (see figure 2.4). Intermediate sizes of ^3J ($\sim$ 2 Hz) coupling values signify conformational averaging, and are therefore not used for inferring torsion restraints.

$^3J_{C_\gamma N}/^3J_{C_\gamma C'}$ are measured by a 2D difference experiment described in section 2.1.4.1. Measurements of couplings to methyl groups are facilitated by the spectral resolution of the methyl ^{1}H,^{13}C chemical shifts as well as the high intensity and small line width of the methyl resonances. Given a experimentally determined pair of $^3J_{C_\gamma N}$ and $^3J_{C_\gamma C'}$ couplings for a methyl group, the identification of its rotermeric state (g^+, t, g^-) is as follows (see figure 2.4):

If a large $^3J_{C_\gamma C'}$ coupling value and a small $^3J_{C_\gamma N}$ coupling value is measured for a methyl

group a of a valine residue, and the inversely situation is found for its intraresidual prochiral partner methyl group b, the conformeric state A is present, and the methyl group a is identified as *pro S* and the methyl group b is assigned as *pro R*. Both couplings have to be available in order to perform stereospecific assignment.

2.2.3 Conformational database potential

A conformational potential term [91] derived from a structure database [95] whose members have been solved crystallographically to high resolution ($\leqq$ 2Å), was implemented into protein model calculation in order to bias backbone ϕ, ψ, and sidechain χ_1 torsion angles to regions of the Ramachandran plot known to be preferentially populated. Because acting directly on rotatable bonds, this energy term presents an useful supplementation to the simple repulsive van der Wall term U_{vdw}.

From the database probability values P_i are extracted by dividing the two dimensional ϕ, ψ space into 2025 8° x 8° bins i (360°/8 = 45° ; 45^2 = 2025), and determining the ratio between the number of examples in the bin (member in the database with ϕ, ψ values within the bin) to the total number of examples in the database. A minimal number of examples in the bin is required to calculate P_i, otherwise examples of the next neighbour bins are also taken into account until a cutoff value is achieved and the new number divided by the number of bins whose data were collected serves as dividend for calculating P_i. Then from statistical thermodynamics, P_i is converted into a potential of mean force by:

$$U_{db}(i) = -k_{db}(\log P_i) \tag{2.42}$$

The weighting factor k_{db} is typically set to rather low values, such that $U_{db}(i)$ is more active on torsion angles of less restrained residues without disturbing the experimentally derived restraints.

In an analog manner, one dimensional energy grids are derived for χ_1 torsion angles.

The conformational database energy term is especially valuable at the early stages of structure of structure determination, where typically only a subset of the total extractable NOE distances can be unambiguously defined, or in regions of the structure model with sparse NOE data.

2.2.4 Radius of gyration for improving the model packing

Structure models derived from short range NMR data tend to be poorly packed and slightly expanded compared with X-ray structures. Although orientational restraints inferred from residual dipolar couplings or paramagnetic restraints, NMR structures are mainly based on short inter-proton NOE distances. Given a network of NOE distances, there are many more possible expanded conformers than tightly packed ones compatible with it. The situation is particular exacerbated when investigating non-globular folds or quaternary structure of a multimere, but also in regions of globular folds were the density of NOE distances is sparse (typically loops, linker regions).

To counteract that entropic effect, a pseudo potential E_{gyr} for the radius of gyrations [92] was incorporated into structure calculation, improving the structural results without affecting or violating other restraints.

The radius of gyration R_{gyr} for a group of atoms is defined as the RMS distance from each atom to their centroid:

$$R_{gyr} = \sum_{j=1}^{N}[r_j - (\sum_{i=1}^{N} r_i/N)]^2/N^{1/2} \tag{2.43}$$

where r_i and r_j are the position vectors of the atoms i and j, and N is the number of atoms.

A good estimate of R_{gyr} for a globular protein or protein domain can be obtained from:

$$R_{gyr} \approx 2.2 \cdot n_{0.38} \tag{2.44}$$

The pseudo potential for N groups of atoms in global environment is incorporated to the target function of a md-SA protocol as:

$$E_{gyr} = k_{gyr} \sum_{n=1}^{N}[R_{gyr-calc}(atoms_n) - R_{gyr-target}(atoms_n)]^2 \tag{2.45}$$

Since there is only one restraint for the selected group of atoms, the weighing factor k_{gyr} (force constant) of the harmonic potential have to be set to rather high values.

R_{gyr} is not a geometrically specific quantity. Restraining a protein structure to a particular target value of R_{gyr} does not restrain it to have any specific fold, but ensures that the structure is overall about as tightly packed as the majority of proteins whose structures have been determined by X-ray crystallography.

2.3 Structure determination of proteins

The goal of a structure calculation is to generate (finally) a real space structure, which satisfy all experimentally and knowledge based restraints. The calculation could be performed either in n-dimensional distance space followed by projection of the obtained self consistent set of distances to into a three dimensional Cartesian coordinate system to get the coordinates of all atoms of the molecule, or directly in real space, which involve minimization of a global target function U_{total} consisting of energy terms for stereochemical and experimental restraints. U_{total} has many false local minima. Therefore, a calculation routine should allow inspection of an extended conformational space without permanent trapping in superficial energy holes. On technique exclusively developed to circumvent such problems and enabling correct protein folding is simulated annealing (SA) [127] using molecular dynamics (*md*).

In molecular dynamics, principles of classical mechanics are employed at the molecular level in order to generate molecular structures in a practical manner as possible. An atom is considered as spherical point mass associated with a net charge, and their interactions are described by classical potentials. In order to evaluate the positions $x(t)$ and velocities $v(t)$ of the atoms at a particular time t, one needs to know the forces $F(t)$ acting on them. For that purpose force fields are derived to approximate the potential energy surface of the biomolecule. So called *Semi-empirical* forced fields were designed by extrapolating thermodynamic and spectroscopic data collected from small molecules to macromolecular level. Thereof, a derivate of the CHARMM forcefield [16] adapted for proteins were used for this work. A target function was minimized via a simulated annealing routine wherein the Newton's equation of motions were solved employing the molecular dynamic program XPLOR [164], as described below.

2.3.1 Molecular dynamic simulated annealing

The target function U_{total} is basing on separable internal coordinates and contains potential terms for experimentally derived restraints U_{exp} and two groups of potential terms for a priori restraints defined by the force field comprising covalent geometry $U_{covalent}$ and pairwise non-bonded U_{non} interactions.

$$U_{total} = U_{covalent} + U_{non} + U_{exp} \tag{2.46}$$

$U_{covalent}$ contains potential terms with internal coordinates for maintaining correct bond lengths,

angels, chirality of all tetrahedral centers and planarity of peptide bonds and aromatic rings:

$$U_{covalent} = \sum_{bonds} k_b(b-b_0)^2 + \sum_{angles} k_\Theta(\Theta-\Theta_0)^2 + \sum_{torsions} k_\varphi(1+\cos(n\varphi-\delta)) + \sum_{improper} k_\xi(\xi-\xi_0)^2 \tag{2.47}$$

whereas b, Θ, φ, and ξ denotes the actual bond length, bond angle, torsion and improper torsion angle, respectively, b_0, Θ_0, and ξ_0 the corresponding equilibrium values, n the periodicity, and δ the phase shift. The parameters k_b, k_Θ, k_φ, and k_ξ designate the weighting factors (force constants).

U_{non} can comprise terms for van der Waals and Lennard-Jones, electrostatic, hydrogen bonding, and solvation interactions. There is a much larger degree of uncertainty in the descriptions and parameterization of those non-bounded interactions. For simplification, only a repulsive van der Waal term was used for the SA scheme of this work:

$$U_{non} = k_{repulsive}(s^2r^2 - r^2) \qquad \text{if } r_{ij} < sr_{min} \tag{2.48}$$

$$U_{non} = 0 \qquad \text{if } r_{ij} \geqq sr_{min} \tag{2.49}$$

where r_{min} are the standard values of the van der Waals radii, s is varied. For supplementation, an empirical database potential for ψ, ϕ, and χ_1 acting rather on rotatable bonds than interatomic distances, were added to U_{non} (see section 2.2.3).

U_{exp} contain all groups of experimentally derived restraints:

$$U_{exp} = U_{NOE} + U_{torsion} + U_{hydrogen_bonds} \tag{2.50}$$

The NOE distance restraints and hydrogen bonds were implemented by a square well potential:

$$U_{NOE} = \begin{cases} k_{NOE}(r_{ij} - r_{ij}^u)^2 & \text{if } r_{ij} \geqq r_{ij}^u \\ 0 & \text{if } r_{ij}^l < r_{ij} < r_{ij}^u \\ k_{NOE}(r_{ij} - r_{ij}^l)^2 & \text{if } r_{ij} \leqq r_{ij}^l \end{cases} \tag{2.51}$$

where k_{NOE} denotes the force constant, r_{ij} the actual distance between two nuclei and r_{ij}^u and r_{ij}^l are the upper and lower bound distance, respectively. In an analog manner, the torsion

angle restraints were formulated by:

$$U_{torsion} = \begin{cases} k_{torsion}(\varphi_{ij} - \varphi_{ij}^{u})^2 & \text{if } \varphi_{ij} \geqq \varphi_{ij}^{u} \\ 0 & \text{if } \varphi_{ij}^{l} < \varphi_{ij} < \varphi_{ij}^{u} \\ k_{torsion}(\varphi_{ij} - \varphi_{ij}^{l})^2 & \text{if } \varphi_{ij} \leqq \varphi_{ij}^{l} \end{cases} \tag{2.52}$$

where φ_{ij} is the torsions angle of the atom i and j connecting bonds, φ_{ij}^{u} and φ_{ij}^{l} are upper and lower bound torsion angles, respectively.

For improving the molecular overall packing, a pseudo-potential for the radius of gyration was incorporated to the target function as given in 2.45.

The time dependent positions and velocities of all N atoms and set of atoms grouped to pseudo atoms (e.g. hydrogen atoms of methyl group or non stereospecific assigned methylene protons) of the molecule were derived using the equations of Newtonian mechanics:

$$m_i \frac{\delta^2 r_i(t)}{\delta t^2} = \vec{F}_i = -\frac{\delta U_{total}}{\delta r_i(t)} \tag{2.53}$$

whereas m_i denotes the atom mass, $r_i(t)$ the position, and $\vec{F}_i$ the force acting on the atom i.

According to the kinetic theory the temperature at a time t is calculated as:

$$T(t) = 2\,\langle \sum_{i=1}^{N} \frac{m_i v_i(t)^2}{2} \rangle / 3Nk \tag{2.54}$$

In principle, initial atomic coordinates at begin of a de novo structure calculation can be any conformation of the macromolecule with good local geometry and no non-bonded contacts (to avoid to large initial forces), as presented, for example, by a extended chain (pre-)minimized employing a few cycles of a simple adopted basis set Newton-Raphson (ABNR) [199] algorithm or POWELL algorithm [18] without consideration of U_{exp}. Disulfide bonds and other covalent connection between sequentially distant residues are generally omitted at the entire stage of a structure determination.

The basic scheme of a molecular dynamic simulated annealing protocol consists of a high temperature phase, where an extended conformational space is available for satisfying the experimental restraints, followed by slow cooling phase in which the structural subtleties are build up.

Initial velocities were assigned to the atoms according to a Maxwell distribution at a defined high temperature. In this work, the temperature was controlled by coupling to a Berendsen heat bath

[110], and the Newtons equations of motions were calculated applying the (very stable) Verlet algorithm [185], in which the position of an atom i at time $t + \Delta t$ is derived from the position of the atom $\vec{r_i}$ at time t by means of a Taylor expansion at $t = 0$:

$$\vec{r_i}(t \pm \Delta t) = \vec{r_i} \pm \frac{dr_i}{dt}|_t \Delta t + \frac{1}{2}\frac{d^2 r_i}{dt^2}|_t \Delta t^2 = \vec{r_i}(t) \pm \vec{v_i}(t)\Delta t + \frac{\vec{F_i(t)}}{2m_i}\Delta t^2 \tag{2.55}$$

Addition of both expansions, $t + \Delta t$ and $t - \Delta t$, leads to:

$$\vec{r_i}(t + \Delta t) = 2\vec{r_i}(t) + \frac{\vec{F_i(t)}}{2m_i}\Delta t - \vec{r_i}(t - \Delta t) \tag{2.56}$$

In that way, the new position of atom i at time $t + \Delta t$ can be inferred from its positions at time t and $t - \Delta t$. The time interval Δt is termed *time step* and must be chosen sufficient shorter than the fastest mode of motion of the simulated system.

A key feature of the molecular dynamic simulated annealing scheme used for this work [127] is the possibility of independent scaling of the weighting factors of all energy terms in equation 2.46. During the initial stage of the high temperature phase all force constants were set to unitary low values, facilitating an enhanced extended conformational space available for finding the global energy minimum minU_{total}. Then, the weightings of the potential terms for maintaining the covalent structure $U_{covalent}$ and of the NMR restraints U_{exp} were successively increased. The repulsion force constant for the repulsive van der Waal term U_{vdW} was leaving initially unchanged, but was chosen sufficient high enough to define and maintain the global structure while allowing atoms to get very close to each other and even move through each other in order to satisfy the restraints. By setting the force constants to very low values at the early stages of the simulation, energy barriers between different global folds can be readily overcome, thus enabling reliable location of global minU_{total}. Before starting the cooling phase the force constants of the repulsive U_{vdW} was also increased to define more and more local structure details. At begin of the cooling phase all force constants have been arrived their final values. In some of the applied md SA protocols, the weighting of the NMR restraints are additionally increased by a factor up to 2 to improve the agreement with the experimental data. Finally, a Powell minimization was performed, but now taking the energy term U_{exp} for experimental restraints into account. Repeating the procedure several times results in a bundle of solution structures whose statistics and quality were assessed by criteria listed in the next section.

2.4 Quality of NMR structures

The quality of a structure can be validated by criteria depending on the experimental input data employed for the structure calculation, but also by a set of several statistical parameters partially derived from empirical databases, which provide a input data independent verification for the correctness of an ensemble of structure models. Both were used as listed below.

2.4.1 Precision of the structure bundle

As measure of precision can be specified either by an average pairwise root mean square deviation (RMSD) of the coordinates from a group of selected models or as an RMSD of the structures relative to the mean coordinates of the ensemble. Both calculations requires the superposition of the structures.

RMSD values can be used to indicate structurally conserved regions within a set of models, but also to identify parts of high structural uncertainty, where additional collecting of experimental NOE data would be necessary.

Another parameter for evaluation of the conformational heterogeneity within a bundle of structure without the requirement of superimposing is the *circular variance* (CV) calculated over a set of ϕ, ψ, χ_1, and χ_2 dihedral angles across all members of the ensemble:

$$\text{Var}(\theta) = 1 - R_{average} \tag{2.57}$$

where $R_{average}$ = R/n, n being the number of members in the ensemble and R for any torsion angle θ (θ=ϕ, ψ, χ_1, or χ_2) is calculated by:

$$R = \sqrt{(\sum_{i=1}^{n} cos\theta_i)^2 + (\sum_{i=1}^{n} sin\theta_i)^2} \tag{2.58}$$

Var(θ) can range between 0 and 1, whereas a low Var(θ) value indicates a tight clustering of the torsions angle values about the mean.

However, RMSD and Var(θ) values a still measures of the precision of a bundle of model structures but not necessarily for their accuracy.

2.4.2 Number and quality of experimental restraints

The precision to which a structure can be derived is directly related to the number of experimental restraints per residue used to generate it. Thereof, the most valuable restraints for determining the protein fold to high resolution are restraints whose related atoms are located far apart along the primary sequence. Furthermore, the extent of stereospecific assignment of prochiral methylene or methyl groups is directly related with obtainable quality of a structure in terms of both, overall packing and precision.

Other criteria for the quality of a structure concerning the completeness of experimentally inferred NOEs, i.e the ratio between the observed distance restraints n_d to the expected ones n_e derived from the atom coordinates of the model, and the agreement with the experimental data, as estimated by a RMSD value:

$$RMSD_{NOE} = \sqrt{\frac{1}{n_d \cdot n_m} \sum_{i=1}^{n_d} \sum_{j=1}^{n_m} \Delta_{ij}} \tag{2.59}$$

where Δ_{ij} denotes the violation of a distance restraint i in the model j to its related upper bound or lower bound, respectively, and n_m is the number of structure ensemble members. An analog expression is calculated for the torsion angles restraints.

An individual verification can performed by listing all restraints with violation that exceeds a cut-off value (typically 0.5 to 1Å for distance restraints and 10° for torsion restraints) or visualizing them on a representative structure model.

2.4.3 Deviation from idealized covalent geometry

The final covalent force constant values related to energy term for bonds, angles, planarity and chirality are set typically set to very high values. Thus, distortions of the local geometry can be addressed to extensive inconsistent distance restraint network and/or bad non-bonded interactions.

2.4.4 Ramachandran maps

Another measure of the quality of the model is obtained by a Ramachandran map, which plots two dimensional distribution of the backbone φ,ψ-torsion angles of a structure. Their empirical combinations in native folds were to be found strongly clustered in preferred areas of the map due to steric repulsion (except for glycine). Because of the pyrrolidine ring, the φ angle of proline

is restricted to an average value around -65°. The structure validation program PROCHECK NMR [96] take use of a two dimensional φ,ψ-torsion probability density distribution constructed from a structure database of 163 structures solved by X-ray crystallography (at resolution about 2.0Å or better), in which four regions (most favoured, additional allowed, generously allowed, and disallowed) were arbitrarily defined according to the density of φ,ψ-torsion angle occurrences within 10°x10° bins of the Ramachandran plot (see figure 2.5.A). During the structure refinement the φ,ψ-angels migrates more and more into allowed conformations, therefore their distribution provide a guide about the quality progress of the subsequently obtained structures. Ideally, over 90% of the φ,ψ-torsion angle pairs located in structured parts of the final model should align within the core regions.

In an analog manner to Ramachandran plots, χ_1, χ_2 scatter plots were produced that indicate the conformity of protein side-chain conformations of a model to known preferences ([96], [107]). Both torsion angles preferentially adopt one of three g_-(≡ clockwise +60°),t, or g_+(≡ -60°) conformers. As refinement proceeds the χ_1, χ_2 torsions becomes more tightly clustered into these three idealized energy wells (see figure 2.5.B). Some statistics of torsion values is given in table 2.1.

Table 2.1: Structure quality indicators used by PROCHECK

Stereochemical parameter	Mean value	Standard deviation
φ,ψ-dihedrals in core regions of Ramachandran plot [%]	$>$90	-
χ_1 torsion angle [°]		
g_-	64.1	15.7
t	183.6	16.8
g_+	-66.7	15.0
χ_2 torsion angle [°]	177.4	18.5
Backbone torsion angles		
Proline ϕ torsion angle [°]	-65.4	11.2
α-helix ϕ torsion angle [°]	-65.3	11.9
α-helix ψ torsion angle [°]	-39.4	11.3
β strand ϕ torsion angle [°]	?	?
β strand ψ torsion angle [°]	?	?
Peptide bond ω torsion angle [°]	180.0	5.8
Virtual ζ torsion angle (C_i^β-N_i-C_i'-C_i^β) for C^α chirality	33.9	3.5

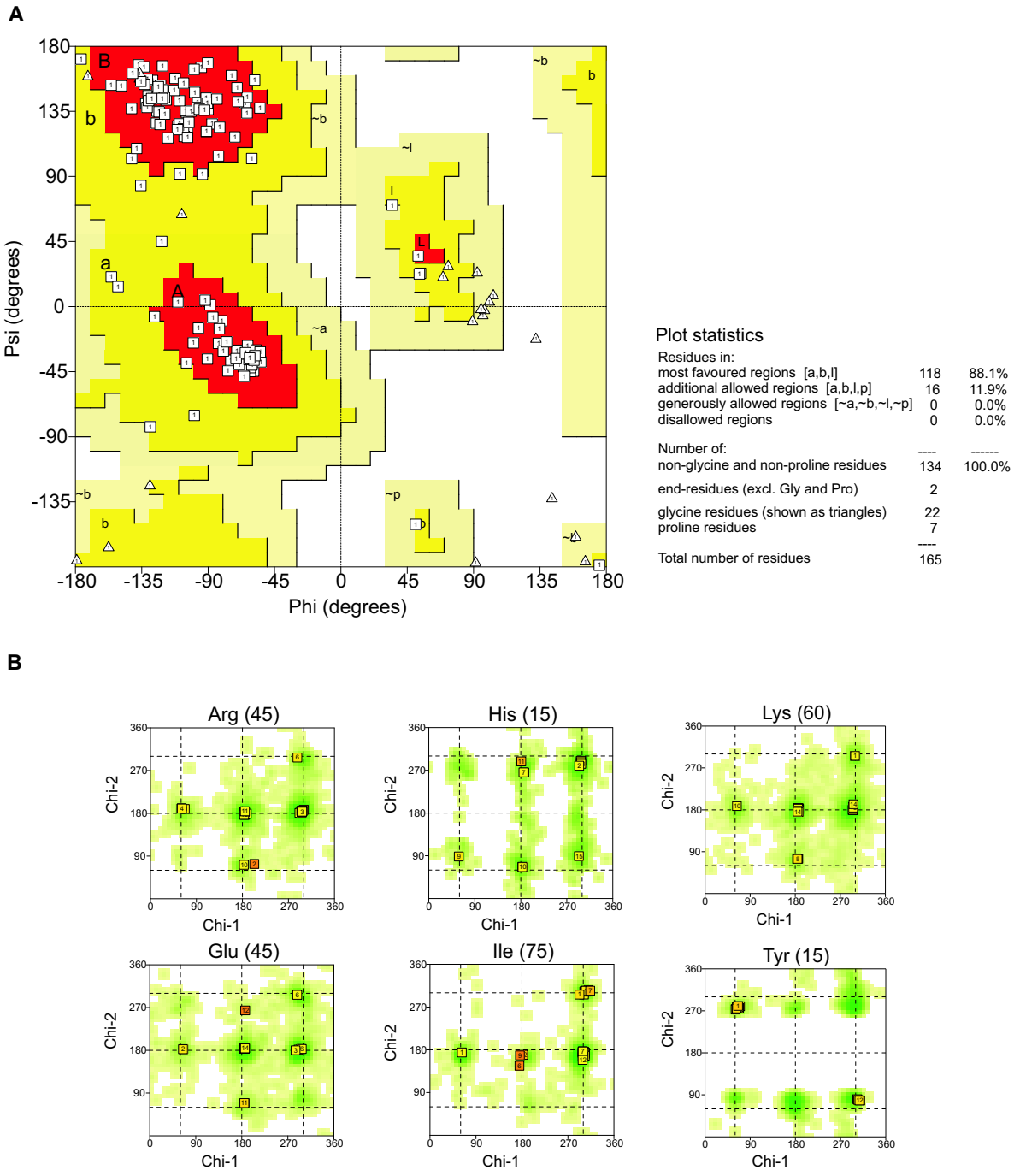


Figure 2.5: (A) Ramachandran and (B) χ_1,χ_2 plot of cyclophilin D. A ϕ,ψ torsion pair of a non-glycine residue in A is indicated by a square whereas those for glycines are represented by triangles (as these are not restricted to the regions of the plot appropriate to the other sidechain types). Degree of dark shading in B states the empirical favorability of a χ_1,χ_2-region.

2.4.5 Contact and environment analysis

Hydrophobic, electrostatic and hydrogen-bonding interactions are the main stabilizing forces of a protein structure. This leads to packing arrangements where hydrophobic residues tend to interact with each other, whereas charged residues tend to be involved in salt links. Hydrophilic residues either prefer to interact with each other or point out into the bulk solvent. Violations of these rules will introduce non-physical interactions, e.g. a charged arginine residue located inside a hydrophobic pocket, that serve as good indicators for model errors. The *PROSA* protein fold validation method [165], [167] make use of inter-atomic energy functions derived from knowledge based mean potentials (which extract information on the forces that contribute to protein folding from a database of experimentally determined protein structures [166]) to characterize the quality of a protein fold.

A method basing on the same principles called *directional atomic contact analysis* [188] validates the overall packing of a protein model by taking also sidechain atoms into account.

Both methods, designed for recognizing misfolded structures and indicating incorrect parts of a given protein conformation, were used throughout that work and therefore summarized below.

2.4.5.1 PROSA

Native folds of proteins are arising from the sum of intra molecular interactions among the atoms of the polypeptide chain, and intermolecular interactions of these atoms with the surrounding solvent molecules. The forces (especially the electrostatic ones) acting on a biomolecule in solution are very complex. While for simulation/description in vacuo the operating forces can be assumed by a simple combination of direct inter-atomic interactions, the situation is completely different in liquid systems, where these direct interactions are heavily disturbed and distorted by the solvent surroundings. The *knowledge based* approach circumvents the problem by extracting the forces and potentials stabilizing native folds in solution from a (large) set of known experimentally derived structures. All inter-atomic interactions that are acting in a structure of a protein (intra- and intermolecular) are described as a function of the two participating amino acids, a and b, the atom types involved c and d, the separation of a and b along the primary amino acid sequence, and the spatial distance r between atoms c and d. Using this model, relative frequencies f_r^{abcdk} for a distinct set of a, b, c, d, and k (which describe the nature of the interaction, e.g. a= Val, b= Ser, c= C^{β}, d= C^{α}, and distance in primary sequence k= 3) from the structure database can be derived, which are a function of the conformation dependent distance

variable r. In the limit of infinitely many observations, these relative frequencies converge to the related probability densities $\lim_{n\to\infty} f_r^{abcdk} \equiv p_r^{abcdk}$. These frequencies are converted to interaction potentials using the inverse Boltzmann relation:

$$E_r^{abcdk} = -kT \ln(f_r^{abcdk}) + kT \ln Z \tag{2.60}$$

where the energy function E_r^{abcdk} is called the *potential of mean force*, and $Z = \sum(-E_r^{abcdk}/kT)$ is the corresponding Boltzmann sum or partition function.

In order to get the interactions or forces specific for a distinct pair of amino acids, the relative frequencies f_r^{cdk}, which contain average statistic of all kind of amino acid pairs a/b, are determined and their related energy potential (corresponding to the energy average over all types of amino acid pairs) is subtracted from E_r^{abcdk} to yield the *netto mean force potential* for a particular pair interaction as follows:

$$\Delta E_r^{abcdk} = E_{abcdk}^r - E_r^{cdk} = -kT \ln(f_r^{abcdk} / \sum_{a,b} f_r^{cdk}) - kT \ln(Z_r^{abcdk}/Z_r^{cdk}) \tag{2.61}$$

ΔE_r^{abcdk} is the energy distribution of the amino acid pair a/b to the averaged energy E_r^{cdk}. E_r^{cdk} is also denoted as the energy frame of reference. By subtracting this reference from the mean force all forces are removed which are common to all subsystems $\sum_{ab} f_r^{cdk}$. The netto mean force contain only those energy components which are specific for a distinct amino acid pair a/b.

Both Boltzmann sums, Z_r^{abcdk} and Z_r^{cdk}, cannot be inferred from the relative frequencies f_r^{abcdk} and f_r^{cdk}, respectively. However, both sums are independent to the conformational variable r, and constant for a particular amino acid sequence. For simplification, it's assumed that $Z_r^{abcdk} \approx Z_r^{cdk}$, thus omitting the second term in equation 2.61. If this isn't valid, the calculated netto energies for a given set of conformations are still shifted by that constant term without effecting the energetic comparison of these different folds.

The applied *PROSA* version uses the individual pair potentials of the 400 natural amino acid combination for the atom types C^α and C^β. Special and effective procedures have been developed and employed by the author to derive useful potentials even for amino acid combination offering a rather small number of observations in the protein structure database [165].

The total intra-molecular pair interaction energy of an amino acid sequence S in any conformation C is derived by summation over all positions i and j in the primary sequence and over all

atom pairs (c,d=C$^{\alpha}$ and/or C$^{\beta}$ in this work).

$$\Delta P(S,C) = \sum_{ij} \sum_{cd} \Delta_r^{a(i),b(j),c,d,k} \tag{2.62}$$

where $a(i)$ and $b(j)$ denotes the amino acid types at sequence position i and j, respectively, and k=$|i-j|$.

Mean force potentials for the intermolecular interaction of the protein atoms with the surrounding solvent molecules can be derived in a way similar to the calculation of the pair interaction mean energy. In the underlying model it's assumed that the complement fraction of the surrounding volume of a protein atom, which is not occupied by other protein atoms contains the solvent molecules. This is expressed by a variable s, which denotes the number of protein atoms in a sphere of a distinct radius R, in which the atom to be determined in respect to its solvent exposure is centered. The netto mean force of solvent exposure of the C$^{\alpha}$ and C$^{\beta}$ atoms belonging to an amino acid type a is obtained by:

$$\Delta E_s^{ac} = -kT \ln(f_s^{ac} / \sum_a f_s^c) \tag{2.63}$$

The total surface energy is computed as:

$$\Delta S(S,C) = \sum_i \Delta E_s^{a(i),c} \tag{2.64}$$

Additionally, one can calculate the combination of pair and surface energy as:

$$\Delta S(S,C) = \omega_p \Delta P(S,C) + \omega_s \Delta S(S,C) \tag{2.65}$$

where ω_p and ω_s are the relative weights of the individual energy contributions.

The quality of an experimentally derived protein fold then is assessed either by its z-score with respect to the energy distribution (of pair interaction and/or surface exposure) over all n possible conformations C or by analyzing the total interaction energy profile of individual residues along the amino acid sequence. The native fold have to be a much lower energy than the average energy over all n accessible conformations, e.g. a large negative z-score. To validate the native state similarity of a structure model, *PROSA* provide a big test set of about 200 alternative conformations from structural unrelated globular proteins, for which the energy distribution is calculated by shifting the query sequence (of the protein to be analyzed)of length l through a

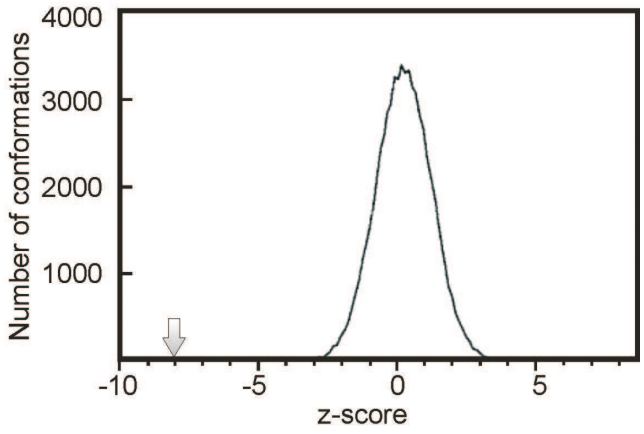


Figure 2.6: Energy distribution of all conformations provided by the artifi cial poly-protein using the lysozyme primary sequence. The arrow indicate the position of the native fold (redrawn from [168]).

given fold of length h of the test set, thus validating against $h - l + 1$ possible conformation.

In order to make results for proteins of different length l comparable, a polyprotein build up by the test set with a total length L about 50000 residues is used by *PROSA* as template for the z-score calculation, so that l is always much lower than L and the number of accessible alternative conformation for a given sequence/fold pair to be validated is approximately the same. The individual protein modules in the polyprotein were carefully bridged by linker regions found in multi domain proteins to avoid steric overlaps or fragments with bad stereochemistry. Although the number of conformations provided by the polyprotein is only a tiny fraction of the huge number of all possible conformations n, it facilitate important regions of the conformation space, because it contain only minimized models of native folds derived from experimental data. Only backbone atoms are present in the polyprotein template.

Starting from the N-terminus of the polyprotein the primary sequence S_p of the protein model to be validated is combined with the first possible fragment of length l, corresponding to conformation C_1 (residues 1 to l of the polyprotein) and the associated conformational energy $\Delta(S_p, C_1)$ is calculated and recorded. Then, the sequence is shifted by one residue towards to the C-terminus and the conformational energy $\Delta(S_p, C_2)$ is stored. This process is repeated for all possible conformations C_q (C_1 to C_{L-l+1}, see figure 2.6).

Finally, the average energy $\bar{E}_p = \sum_q \Delta E(S_p, C_q)$ and its associated standard deviation σ_p are computed. Then, the energies sampled along the polyprotein are transformed to z-scores:

$$z_{p,q} = (\Delta E_{S_p,C_q} - \bar{E}_p)/\sigma_p \tag{2.66}$$

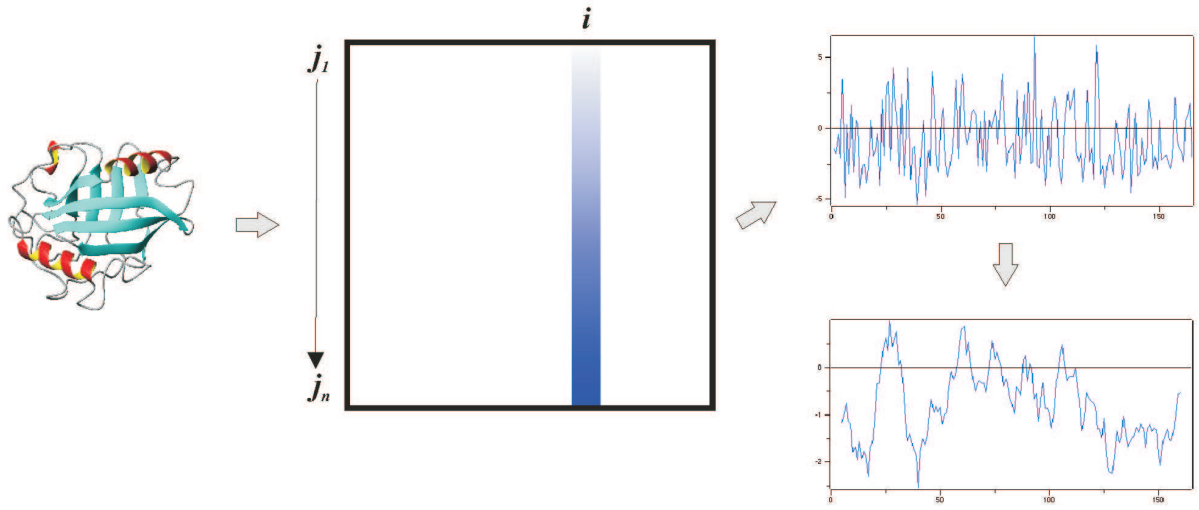


Figure 2.7: Total interaction energy as a function of primary sequence position. The energy fluctuation of the initially derived profile (plot A) is usually damped (for a better representation) using a moving window average over a set of residues (a window size of 10 residues was applied for generating plot B)

An example is given in figure 2.6.

A detailed view about the energy distribution within a conformation is obtained from the total pair interaction energies of the each residue along the amino acid sequence. The total pair interaction of a residue at sequence position i with respect to all other residues obtained is:

$$E_{i,total} = \sum_{j} e_{ij} \tag{2.67}$$

which corresponds to the sum over a column i of the pair energy matrix e_{ij} (figure 2.7). The resulting plot serves as a useful energetic fingerprint. For native folds, the energy is below zero for the most sequence positions and only occasionally small positive peaks are observed. Large positive peaks indicate faulty parts of the chain. An analog diagnostic tool is generated by using the surface energy (according equation 2.64) calculated for each residues as ordinate. The individual energy values of the respective residues are represented in E/kT units, where k is the Boltzmann constant and T is the absolute temperature (kT = 0.582 kcal/mol for T = 295K). For better visualization, the (strongly fluctuating) energy profiles are damped using gliding averages along the amino acid chain.

2.4.5.2 Directional atomic contact analysis

Another approach (using empirically derived probability distributions) for verifying the proper packing of any protein model calculates a contact quality index, which is a measure of the

agreement between the probability distributions of atoms around each residue fragment in the model and equivalent distributions derived from a database [188]. In contrast to PROSA, this quality index is evaluated for both heavy backbone and sidechain atoms. The fragments are defined so that they are as large as possible, but do not contain rotational degrees of freedom around torsion angles. To increase the number of atomic contacts as far as possible for statistical relevance, the 167 atoms types of the natural 20 amino acids were grouped into 57 atomic classes. For instance, all backbone nitrogen's atoms except those for glycine and proline (forming separate classes) represent one class of atom types.

From a database of 60 x-ray models (resolution leq 2.5Å) contact probability densities for each atom class around each fragment type are calculated on equably spaced grid points (1Å intervals) arranged in a cubic box around each fragment. The box size (16Å x16Å x16Å) were chosen to enable accommodating of the largest fragment (the tryptophan double ring system). For each fragment grid point (centered in the grid box) in a protein of the database a value is added whenever the distance to an atom i of an fragment contacting residue is shorter than the van der Waal radius of i. A fragment- surrounding residue contact is assumed if at least one interatomic distance is shorter than the sum of the distance related atoms plus the radius of the water molecule ($\sim$1.5Å). The individual added grid point value is determined by:

$$V(d) = \quad 1 - d^2/R^2 \quad \text{if } d < R$$
$$V(d) = \qquad 0 \qquad\quad \text{if } d \geq R$$

Subsequent the accumulated grid point values are normalized by the frequency of occurrence of the fragment type in the protein database. Finally two sets of probability distributions ($P(s_{helical},fr,atm,x)$ and $P(s_{non-helical},fr,atm,x)$) were calculated by summation over all fragment (fr) and atom (atm) types belonging to helical or non-helical secondary structure s, whereas x denotes a distinct grid point of the box. P(s_x,fr,atm,x)

To obtain a significant measure for the packing quality of a model only contacts high frequent in the database are taken in such test into account. For each fragment fr' of the model a quality index is evaluated by

$$Q(fr') = \sum_{atm'} \sum_{x} A(fr')V(d)P(s_x, fr, atm, x) \tag{2.68}$$

where atm' is the atom type in the fragment fr' contact making residue, $V(d)$ (which is a func-

tion of the atom type and the grid point of the fragment surrounding box) is given by equation , $A(fr')$ is the number of unique atoms in the fragment fr', and $P(s_x,fr,atm,x)$ is the probability density for atom type atm' in the environment of fragment type fr' at grid position x. The fragment fr' resides in a residue with secondary structure s_x (x=helical or non-helical).
Then these $Q(fr')$ values are summed over all fragments in the residue (or residue range).
For an absolute scale of the derived
To obtain an absolute scale for the Q-values, average quality indices $\bar{Q}$ were determined for any particular residue type from the database as a function of secondary structure type and surface contact area (residues at the surface make fewer contacts than residues of the same type in the core). Any computed residue quality index Q is then related to the average quality index $\bar{Q}$ of the same residue type and the same surface contact area, and transformed to z-scores:

$$z = \frac{Q - \bar{Q}}{\sigma} \tag{2.69}$$

were σ denotes the standard deviation obtained from the database.
A z-score lower than zero indicates a more unusual residue environment than found for the average. In contrast, a positive z-score shows a more typical surrounding for that residues as found for the mean. An obtained z-score for a residue lower than 5 indicate that there is something wrong or an involvement of that residue in intermolecular contacts. Faulty parts or the overall quality of a model fold can be assessed from overall quality control z-scores (averaged over a particular range of amino acid sequence). Very low z-scores for a group of sequentially connected residues suggest a potential wrong arrangement of this protein part.
From comparing a database of well refined proteins with a set of misfolded protein structures a classification for that overall quality control z-scores was tabulated (see table 2.2).

Table 2.2: Empirical interpretation of the quality index for NMR structures obtained by DACA.

overall quality control z-score	classification
below -5.0	fold most likely incorrect
-2.0 to -5.0	poorly refined structure or structure containing miss-arranged parts
above -2.0	structure with good overall packing

2.5 Towards automated structure generation

2.5.1 Automated backbone assignment using the MONTE program

An automated Monte Carlo based approach for backbone chemical shift assignment has been used to test the effect of several low resolution structures (offering a broad range of different fold topology) obtained by a semi-automatic direct method for fast structure evaluation in terms of assignment quality. In addition, the backbone assignment of Cyclophilin D was facilitated by a similar program developed by [65].

Although of robust assignment program has been already suggested earlier, Monte Carlo based programs are particular useful because they explore the landscape of possible solutions during the assignment process. Furthermore, they are especially robust against missing data, e.g. absent backbone amide $^{1}\mathrm{H}^{N}$ resonances. By beginning from various distinct starting points, also the reliability of an assignment proposal can be derived from the result statistics.

Generally such automated methods work especially efficient if the major part of the spin systems can be defined unambiguously.

The MONTE program drys to find the best global mapping of the spin-systems onto the primary sequence by using a standard Metropolis Algorithm [117] in connection with simulated annealing. The routine is executed for several numbers of cycles to generate an ensemble of possible solution, from which the reliability of the individual assignments can be assessed.

Each cycle starts with a random mapping of the experimentally derived spin systems to a residue position of the primary sequence or within a buffer, whose size is 20% of the protein length. The buffer facilitates temporary removal of spin systems from the scoring process, permitting Monte Carlo moves that might otherwise unlikely because of poor score. Spin systems arising from artefacts or impurities will reside in the buffer at the end of the assignment process. The initial assignment is scored over all spin systems mapped on the primary sequence (excluding spin systems in the buffer) by a function that can contain any combination of the following terms:

- Inter residue chemical shift matching If two spin systems are mapped to adjacent residues on the primary sequence, the program compare the inter residue chemical shifts associated with spin system i with the intra chemical shifts of the spin system assigned to residue $i-1$. The difference of these chemical shifts $\Delta\delta$ are used as argument to a Gaussian function, at which the user can define a cutoff level value below $\Delta\delta$ give rise

for a negative (e.g. unlikely) scoring. This provides a repulsive terms for detaining of mismatched inter residue chemical shifts.

- Correlation of experimentally derived backbone amide $^{1}H^{N}$-$^{1}H^{N}$ NOEs with the NOE pattern predicted from a provided tertiary structure Each difference of the amide chemical shifts between proton atoms that are currently participating in the NOE are evaluated (as below by a Gaussian scoring function) and summed as contribution for the total score. Alternatively, if only a secondary structure is known, the program generates a structure template with α and β segments for scoring. Where no structure information is available a β strand conformation is created to predict sequential NOEs.

- Conformability of the intra i and inter $i-1$ residue chemical shifts of a spin system with the empirical chemical shifts of the residue types (of position i and $i-1$) to which it is mapped. For that residue type matching three sets of chemical shift distributions extracted from the BioMagResBank(BMRB) for amide nitrogen N, C^{α}, C^{β}, and carbonyl C' atoms are employed in this work. Each set contains chemical shift values associated with α-, β- or random coil conformation.

- Residue type information from specific labeling Using of ^{15}N labeled amino acids [114] or methods of selective methyl group protonation of perdeuterated proteins [152], [182], or residue specific pulse sequences [162] permit the identification of the amino acid type of a spin system. By selective ^{13}C' labeling of particular amino acid type in an otherwise uniformly ^{15}N labeled sample residue type information of the residues preceding that carbonyl position can be extracted by employing the $^{1}J_{NC'}$ coupling for an ^{13}C-edited HSQC spectrum. This information can also be provided as input for a MONTE run.

After the initial random assignments are scored, a simulated annealing Monte Carlo routine dry to optimize the mapping of the spin systems to the primary sequence by exchanging, one or more consecutive spin systems from within the primary sequence with an identically sized collection of spin systems from either the primary sequence of the buffer. These segments are selected randomly from the primary sequence and are of randomly size, with a maximal length defined by the user. Subsequent, the new mapping is scored. If the new score S_{new} is higher (better) than the previous one S_{old}, than the mapping is accepted. If not, the new arrangement is accepted only if a probability factor P ($= \exp[-(S_{new} - S_{old})/T]$) basing on the ratio of the decrease in respect to the current SA temperature T is greater than a random

number p (ranging from 0 and 1). The smaller the difference $\Delta S = S_{new} - S_{old}$ and as higher the simulation temperature T, the probability for accepting the new mapping increase. The parameter T is set to high initial values such that the most proposed changes are retained. Subsequent the temperature is gradually lowered during the run, consequently it becomes increasingly less likely to accept changes that increase the overall score. A large number of randomly proposed changes are necessary (in the order of 10^4) at each time step to ensure that the systems remains in equilibrium. Beside output text files the program generates nice summary output files in postscript format for quick examination of the obtained results (see figure 2.8).

In figure 2.8.A the reliability of the assignment is indicated by (uncertainty) bars (first line top), whose highs are proportional to the number of different spin systems that were assigned to their particular residue positions in independent Monte Carlo assignment runs. Open rectangles mark the location of Pro residues. A zero height bar (thin line, low uncertainty) indicates that consistently the same solution was obtained. The next line show the information derived from the input NOE data (optionally) in context with user provided structure model. The rows labeled with J_x indicate the fidelity of matching of inter- and intra-residue chemical shifts, whereas the thickness of the lines is related to how well the chemical shifts match. At the top the input primary sequence is given. In 2.8.B a correlation plot indicating the uniqueness of assignments is given. Most frequently obtained sequential assignment is given in the diagonal. Alternative assignments are printed outer diagonal in the plot. The intensity of the plotted points are proportional to the frequency that a spin-system is assigned to a particular residue (more darkly colored refer to a higher the frequency a single spin-system to a particular residue). For example, a stretch of 21 connected spin-system were alternatively assigned by MONTE to residue position Phe41 to Cys62 (arrows in magenta), whereas most frequently this spinfragment was assigned to residue Ser7 to residue Ser28. Note that only triple resonance data of $^{13}C^{\alpha}$ chemical shifts were part of the MONTE input in that particular case.[1]. See MONTE manual for more information.

[1] resulting in a low discrimination power between different residue position even for such a long stretch of connected spinsystem

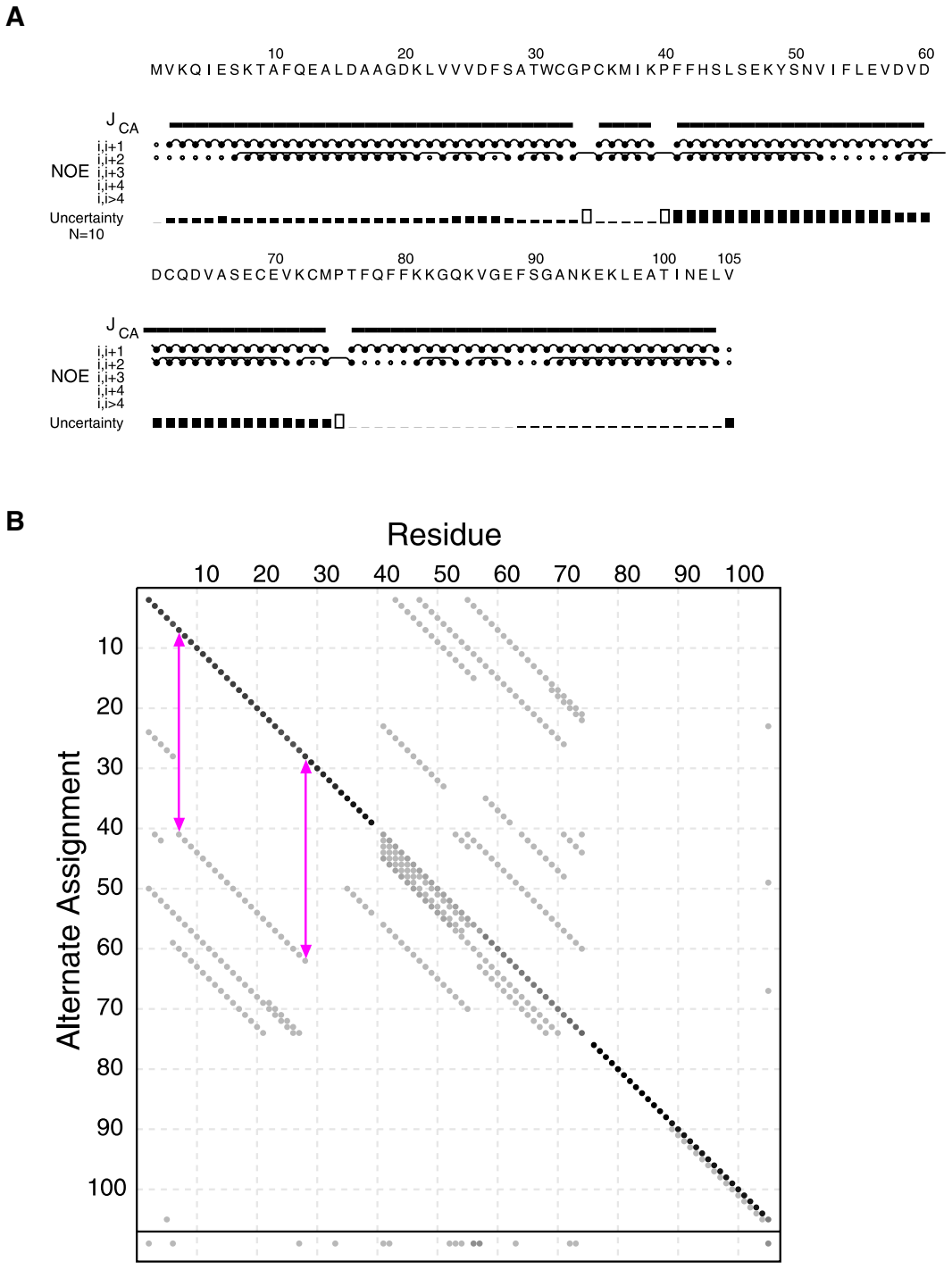


Figure 2.8: (A) Graphical summary of the fi nal assignment results generated by MONTE[63]. (B) Correlation plot indicating the existence of alternative assignments. See text for more details.

2.5.2 Simplification and sensitivity enhancement of NMR spectra using protein deuteration and specific labeling patterns

Sensitivity of ^{15}N, ^{13}C triple resonance NMR experiments employed for assignment of protein backbone and sidechain chemical shift resonances is affected adversely by increases in the transverse relaxation rate constants for ^{15}N and ^{13}C spins. Whereas ^{13}C relaxation is predominated by the ^{1}H-^{13}C dipolar interaction, the relaxation of the backbone $^{1}H^{N}$ spins arises substantially from dipolar interaction with proximal aliphatic proton spins. However, eliminating of these specific relaxations pathways by replacing most or even all carbon bound hydrogen atoms by deuterium ^{2}H can significant increase both spectral resolution and sensitivity of NMR spectra.

Furthermore, (fractional) protein deuteration and the use of specific labeling (see below) considerable simplify the interpretation of NOESY spectra by decreasing spectral overlap and line broadening, thus allowing extraction of a higher number of NOEs and definition of an extended magnitude of initial unambiguous NOE distance restraints, both a prerequisite for obtaining a three dimensional protein structures of reasonable accuracy and precession. Because deuteration limits spin diffusion by reducing the number of available magnetization transfer pathways, it's possible to infer internuclear distance from cross peak intensities more accurately when using longer mixing times.

Complete deuteration, or perdeuteration, replaces about 99% of all carbon bound protons. In contrast, via random fractional deuteration only a destinct percentage (up to 80%) of these hydrogen atoms are exchanged (randomly distributed over all molecules in the sample - with exception of acidic carbons) according the choosen D_2O/H^2O ratio in the minimal media culture for expressing the protein. Perdeutered protein samples can be obtained using deuterated sources (e.g. ^{2}H-glucuse or ^{2}H succinate) or simple protonated carbon sources (e.g. ^{1}H sodium acetate) were the carbon bond protons are replaced by solvent deuterons prior to or during amino acid biosynthesis [152].

However, perdeuteration diminish both the maximal number and the nature of possible proton-proton NOEs.

Additional NMR information in deuterated protein can readily be obtained from NOEs of side chain amide positions of argenine, asparagine, and glutamine, which can in principle, be fully protonated by dissolving the deuterated protein in H^2O.

Useful labeling patterns arises from the application of protonated carbon compounds as growth

substrate whose protons are effectively retained at specific sites at several amino acids. [For example, the use of ^{1}H-glucose, whose carbohydrate derivates serves as precursor in biosynthesis of aromatic groups, give rise for a high protonation level in aromatic sidechains.]

More recently developed labeling methods are concerned with selective protonation. The goal of such approaches is to introduce specific labeling patterns in order to simplify the spectra without losing requisite structure information. All this techniques involves selective reintroducing of ^{1}H spins into non-exchangeable sites of otherwise perdeuterated proteins. Successful selective protonation of aliphatic methyl groups of leucine, isoleucine (H^{gamma2} only) and valine residues was reported the first time by Rosen and Co-workers [152]. They use protonated pyruvate (as the sole carbon source), which is the metabolic precursor of that methyl groups, for protein expression in a $^{2}H_2O$-based minimal media. However, using of pyruvate instead of glucose lead to a about two fold reduced yield of expressed protein.

Because these methyl groups are preferentially located at the hydrophobic protein core or at the interface of multi-protein complexes, very useful long range NOEs for defining a tertiary fold or biomolecular arrangement can be collected from them. Furthermore, methyl groups can serve as excellent reporters of protein dynamics and provide insights into functionally relevant protein motions [77]. Although these methyl functions show a somewhat limited chemical shift dispersion, their usual long relaxation T_2 times (due to rapid rotation about the methyl symmetric axis) give rise for sharp and intense signals, even in ^{13}C labeled samples.

A serial of methods [182],[101] use the biogenic precursors ^{13}C 2-ketobutyrate and ^{13}C 2-ketoisovalerate for selective protonation of leucine H$^{\delta}$, valine H$^{\gamma}$ and isoleucine H$^{\delta 1}$ methyl functions. Beside high level ($>$90%) protonation level an additional benefit arise from the absence of different isotopomers CH_iD_j, which otherwise would result in a distribution of ^{13}C aliphatic chemical shift (thus lowering both sensitivity and resolution of the spectra) especially for methylene and methyl groups caused by ^{2}H isotopic effects. The ^{13}C chemical shift of an aliphatic ^{13}C carbon not only depends on the isotopomeric state (i and j), but also on the isotopomeric states of aliphatic carbon groups one and two bonds apart. The second methyl group of valine and leucine residues remain unlabeled and is deuterated (when using $^{2}H_2O$ growth medium) in order to observe valueable inter-methyl NOEs only. Another advantage of that methods is the protonation of the H^{delta1} position (rather than H^{gamma2} position), because the ^{1}H, ^{13}C chemical shift of these methyl groups are often significantly better resolved than those from other methyl functions, even in very large proteins.

In contrast to the protocols of [182] and [152] the concept for the synthesis of the ^{13}C 2-

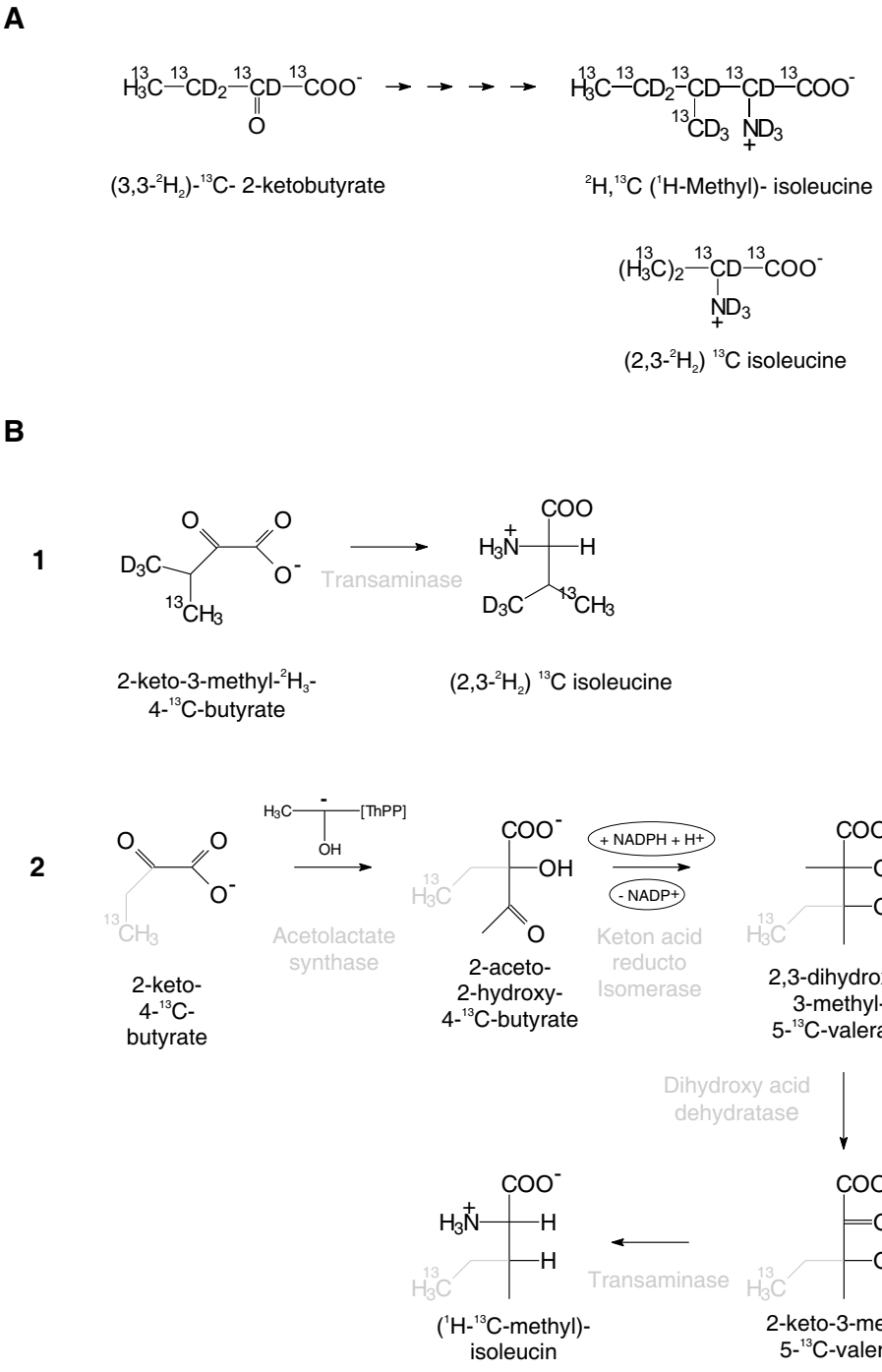


Figure 2.9: Selective labeling scheme for residue types leucine, isoleucine and valine according (A) [182] and (B) [101].

ketobutyrate and ^{13}C 2-ketoisovalerate precursors suggested by [101] allows selective ^{13}C labeling of the terminal methyl carbon only (instead of uniformly ^{13}C labeling) of that precursor compounds. Furthermore, except for the second ^{13}C methyl group, the side chain carbons of the valine, isoleucine and leucine residues remain protonated, thus allowing detection of NOEs between methyl $^{13}CH_3$, amide $^1H^N$, H^α, and sidechain protons (see figure 2.9 for comparison). Useful NMR experiments for the easy assignment of leucine H^δ, valine H^γ and isoleucine $H^{\delta 1}$ methyl ^{1}H, ^{13}C chemical shift in samples generated starting from ^{13}C labeled precursors were presented for test cases as big as malate synthase G (723 residues) by Kay and Co-workers. Measurement of valuable unambiguous $^1H^N$-$^1H^N$, $^1H^N$-$^{13}CH_3$, and $^{13}CH_3$-$^{13}CH_3$ NOEs from samples described above is facilitated for example from 4D ^{15}N-^{15}N, ^{15}N-^{13}C, and ^{13}C-^{13}C edited NOESY spectra, respectively. In favorable circumstances, both an uniform labeled ^{15}N, ^{13}C sample and a selective labeled sample should be available to extract a high content of unambiguous NOEs for structure determination. If only a selective protonated sample is existent, their limited NOE content (because of deuteration) could be supplement by residual dipolar couplings and anisotropic chemical shift changes measured in weakly aligned samples [7].

In addition and of particular interest is the selective protonation of aromatic sidechains of phenylalanine, tyrosine, and tryptophane residues, because they are also prevalent at hydrophobic cores. A biosynthetic approach has been suggested by [145] in which specific protonation of aromatic sidechains with complete deuteration at C^α/C^β position was accomplished in proteins over-expressed in bacteria grown in shikimate[2]-supplemented ^2H_2O medium. The protonation of backbone H^α and H^β sidechain positions can be controlled independently via the ^2H_2O and addition of glucose to the growth media.

More recently, a highly promising approach termed SAIL (stereo-array isotope labeling) was introduced by Kainosho and Co-workers [70]. This method provide stereo and regiospecific labeling patterns for all 20 natural amino acids prepared by chemical and enzymatic syntheses. This includes stereoselelctive replacement of one ^{1}H atom by ^{2}H, two ^{1}H protons in methyl groups by ^{2}H, stereoselective labeling of the isopropyl groups of valine and leucine residues as given used by methods mentioned above, and in addition labeling of six-membered aromatic rings by alternating ^{13}C-^{2}H and 13C-^{2}H moieties.

[2]Shikimic acid is a precursor to the aromatic phenylalanine, tyrosine, and tryptophane residue types.

2.5.3 Protocol for automated methyl group assignment

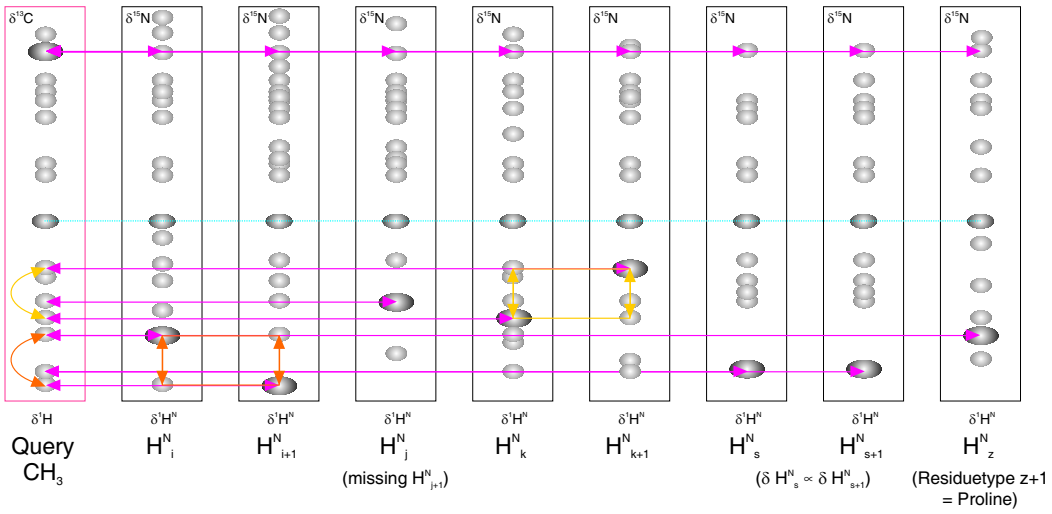
Once the assignment of the backbone chemical shifts and a set of unambiguous backbone amide $^{1}H^{N}$-$^{1}H^{N}$, $^{1}H^{N}$-$^{13}CH_3$(ILV), and $^{13}CH_3$(ILV)-$^{13}CH_3$(ILV) (where CH_3(ILV) denotes protons of leucine H^{δ}, valine H^{γ} and isoleucine $H^{\delta 1}$) NOEs are available, the assignment of the sidechain methyl groups allows conventional structure calculation of a convenient initial fold model. Subsequently such models can be successfully and comfortably refined by automated approaches described in chapter 1.

A protocol (written in *perl* language) for automated methyl group assignment developed and tested in this work (figure 2.1) uses as input the primary sequence of the protein, a list containing the ^{1}H,^{13}C chemical shift of the unassigned methyl groups, and a list encompassing NOE cross peak chemical shift coordinates of CH_3-X NOEs (where X can be any arbitrary proton) extracted from 3D ^{15}N- and ^{13}C-edited NOESY spectra, respectively, for a test set of three proteins, Cyclophilin D, ICln, and LIM1(CRP2), that offers different fold topology, extent of accessible chemical shifts, chemical shift dispersions, and molecular sizes. The backbone and sidechain chemical shifts of these proteins were previously manually assigned in this work from a ^{15}N, ^{13}C uniform labeled sample. For simulation purposes, the chemical shift coordinates of the methyl groups (but not any residue type information) for the program input were assumed to be known from employing a selective protonated sample. The NOE database used for testing the procedure were exclusively generated from NOESY peak lists available from those conventional assignments and contained (predominantly ambiguous) entries for *unassigned* methyl groups of residues leucine ($CH_3^{\delta 1,2}$), isoleucine ($CH_3^{\gamma 2}$,$CH_3^{\delta 1}$), valine ($CH_3^{\gamma 1,2}$), alanine (CH_3^{β}), and threonine ($CH_3^{\gamma 2}$), and assigned backbone amides $^{1}H^{N}$ of the form X-$^{1}H_i \leftrightarrow {}^{1}H_j$, where X denotes the heteronuclei ^{15}N or ^{13}C, respectively.

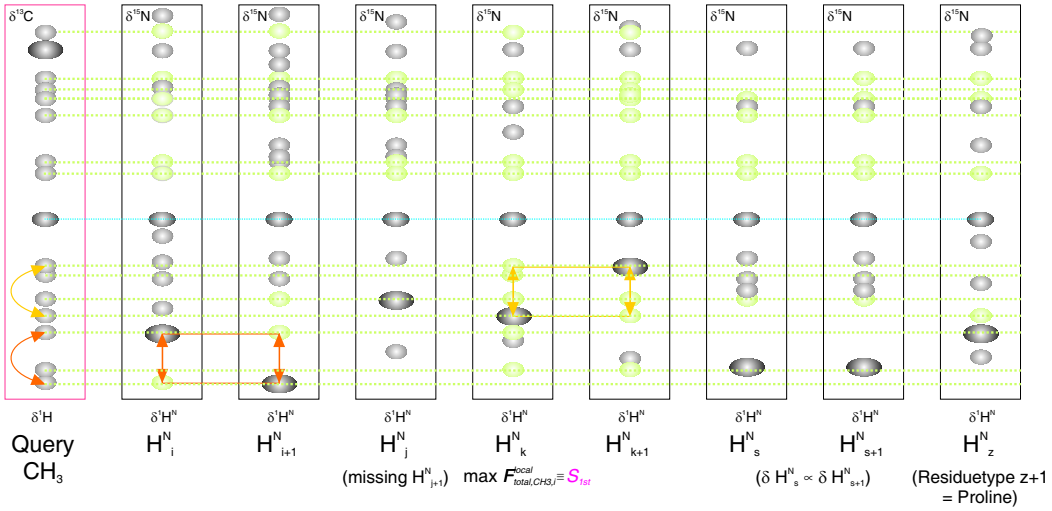
Because of their experimental origin, all lists of shift coordinates used here are associated with uncertainties due to incomplete picking (because of inexistent signals), noise, and pick errors in crowded spectral regions. However, special care was taken that the lists didn't contain peaks of artifacts arising from spectral processing or sample impurities. Elaborative manual inspection was performed to resolve NOE traces in the 3D ^{15}N- and ^{13}C-edited NOESY spectra of degenerated $^{1}H^{N}$/^{15}N or ^{1}H/^{13}C chemical shift pairs, respectively.

The program starts with a chemical shifted based (peak symmetry related) correlation of a methyl group with ^{1}H/^{13}C chemical shift coordinates to all matching (assigned) backbone amide $^{1}H^{N}$ protons, i.e. a list for an unassigned methyl group i with chemical shifts $^{1}H_i$/$^{13}C_i$ is created

A Chemical shift and sequential CH_3-H^N_{resnr}-$H^N_{resnr+1}$ NOE correlation based initial selection



B Common CH_3-H^N NOE pattern analysis



C Scoring of the assignment possibilities and local classification as unambiguous, ambiguous, and potentially wrong.

$F^{local}_{total,CH3,i}=F_{match}+F_{NOEpattern}+F_{emp}$

$p_{21}=S_{2nd}/S_{1st} < p_{cl}$ && $S_{1st} > p_{min}$?

```
-------------------------------------------------

manual assignment: 78.ILE.HD11

assignment possibilities found:

S1st  78.ILE.HD11  36.5   -> unambiguous assigned
      78.ILE.HG21  31.5
S2nd  38.ALA.HB1   10.6

-------------------------------------------------
```

D Global classification of the local assignments

Locally proposed potential candidates C_x elsewhere unambiguously assigned ?

Recalculated $p_{21,new} < p_{cl}$ && $S_{1st} > p_{min}$?

```
-------------------------------------------------

manual assignment: 139.VAL.HG21

assignment possibilities found:

S1st 139.VAL.HG#   17.0   -> ambiguous assignment
C1   142.ILE.HG21  13.7 1 -> already unambiguous assigned
C2   142.ILE.HD11  12.8 1 -> already unambiguous assigned

-------------------------------------------------
```

Figure 2.10: Flowchart of the applied routine for automated assignment of methyl groups.

containing all $\sum_{j=1}^{N}$ $^{1}H_{j}^{N}$ amide functions for which a symmetrical NOE pair $^{13}C_i$-$^{1}H_i$ $\leftrightarrow$$^{1}H_{j}^{N}$ and $^{15}N_j$-$^{1}H_j$ $\leftrightarrow$$^{1}H_i$ in the ^{15}N- and ^{13}C-edited NOESY spectra exists, respectively. These entries can be scored as F_{match} according the difference in their proton chemical shifts within user defined cutoff values for shift matching using a Gaussian function.

Then, the part of NOEs common to both the methyl group i and to a backbone amide $^{1}H_{j}^{N}$ is quantified and normalized to the total NOE number of methyl group i and the backbone amide.

$$F_{NOE_pattern,i} = \#NOE_{i,j}/\#NOE_i \tag{2.70}$$

Subsequently, all backbone amide $^{1}H^{N}$ functions adjacent in the primary sequence are evaluated and listed, whose proton chemical shifts are found in the ^{1}H/^{13}C NOE pattern of methyl group i. That criteria makes use of the fact, that independent of the secondary structure at least *intra*-residue H_i^N-$CH_{3,i}$ NOE contacts and NOEs between a methyl group $CH_{3,i}$ to the backbone amide H_{i+1}^N consecutive following in the primary sequence are observed when using common NOESY mixing times τ_{mix} (e.g. 150ms or more). Beside entries of the form H_i^N/H_{i+1}^N this list for a methyl group $CH_{3,i}$ additionally contains entries of single H_i^N atoms, whenever the residue type of the amino acid following in the primary sequence is a proline, or an assigned H_{i+1}^N is still missing or show approximatively the same $^{1}H^{N}$ chemical shift. The criteria is particular effective if both amide protons, H_i^N and H_{i+1}^N, exhibit different $^{1}H^{N}$ chemical shifts (as normally present).

Table 2.3: ^{1}H and ^{13}C methyl chemical shift statistics extracted from the BioMagResBank.

Residue type	Group	δ^{1}H [ppm]	δ^{13}C [ppm]
Alanine	CH_3^{β}	1.37 ± 0.28	19.0 ± 2.8
Isoleucine	$CH_3^{\gamma 2}$	0.75 ± 0.51	17.7 ± 3.7
	$CH_3^{\delta 1}$	0.64 ± 0.57	13.8 ± 4.1
Leucine	$CH_3^{\delta 1,2}$	0.75 ± 0.44	24.5 ± 2.7
Methionine	$CH_3^{\epsilon 1}$	0.43 ± 0.53	17.4 ± 4.5
Threonine	$CH_3^{\gamma 1}$	1.16 ± 0.55	21.7 ± 3.7
Valine	$CH_3^{\gamma 1,2}$	0.80 ± 0.60	21.5 ± 2.7

Afterwards, a score $F_{total,CH_{3,i}}^{local}$ for each H_i^N-$CH_{3,i}$-H_{i+1}^N fragment is evaluated and stored as local assignment of $CH_{3,i}$:

$$F_{total,CH_{3,i}}^{local} = \omega_{match} \cdot F_{match} + \omega_{NOEpattern} \cdot 1/N \sum_{n=1}^{N} F_{NOEpattern,n} \tag{2.71}$$

where ω_{match} and $\omega_{NOE_pattern}$ are weighting factors, respectively, and N denotes the number of atoms (1 or 2, corresponding to single H_i^N atom or H_i^N/H_{i+1}^N atom pair) in entry i.

This score can additionally also include a probability term F_{emp} for empirical chemical shifts δ_{emp} extracted from the BioMagResBank (BMRB, http://www.bmrb.wisc.edu/; see table 2.3).

The contribution of F_{emp} to the total score depends on the differences of the experimentally observed and empirical shifts $\Delta\delta = |\delta_{obs} - \delta_{emp}|$ and the standard deviations σ_{emp} in form of a Gaussian function:

$$F_{emp} = \sqrt{\frac{1}{\pi\sigma_{emp}}} \cdot e^{-\Delta\delta/\sigma_{emp}} \tag{2.72}$$

However, the term F_{emp} is not applied for assignment of methyl groups with conspicuously ^{1}H/^{13}C chemical shifts (e.g. due to spatial proximity to aromatic sidechains). Such methyl groups are generally found towards the protein core, where they typically form a broad set of valuable NOEs (with characteristically chemical shifts) and therefore can be assigned easily without using F_{emp}.

Although not used here, however, once available a primarily structure could be implemented to further restrain the assignment possibilities.

The local assignment scores $F_{total,CH_{3,i}}^{local}$ for a methyl group i are ranked and reported. A local assignment is assumed to be unambiguous whenever a best placed assignment score $S_{1^{st}}$ exceeds a user defined threshold and the probability ratio $p_{21} = S_{2^{nd}}/S_{1^{st}}$ is lower than a cutoff p_{cf}. Finally, a completing global assignment is performed by mutual exclusion of the individual unambiguous local assigned methyl groups in ambiguously assignments. Additionally, a criteria can be defined demanding for a maximal agreement in the NOE patterns of the individual assigned prochiral methyl groups of residue type valine ($CH_3^{\gamma 1}$ and $CH_3^{\gamma 2}$) and leucine ($CH_3^{\delta 1}$ and $CH_3^{\delta 2}$), respectively.

2.6 Semi-automated protein structure calculation with a priori sequential backbone resonance assignment

Once a list of spinsystems preferentially containing at least two sets of carbon atom types (e.g. $^{13}C^{\alpha}$/$^{13}C^{\beta}$ or $^{13}C^{\alpha}$/^{13}C') can be unambiguously defined the sequential backbone chemical shift assignment for a protein can be reliable obtained applying a Monte Carlo approach [63] as described in section 2.5.1 or by an analogous method [65]. The use of selective labeling techniques presented in section 2.5.2 in context with an automated routine for methyl group

assignment facilitate the generation of valuable assigned CH_3-H^N and CH_3-CH_3 NOE distance restraints for structure determination.

Then, the input data can principally contain:

- A set of H^N-H^N, CH_3-H^N and CH_3-CH_3 NOE distances Unambiguous H^N-H^N can be derived from 4D ^{15}N/^{15}N edited NOESY experiments, CH_3-H^N and CH_3-CH_3 NOE are extractable from 3D ^{15}N- and ^{13}C-edited (provided a favourable shift dispersion) or from 4D ^{13}C/^{15}N and ^{13}C/^{13}C-edited NOESY spectra, respectively. To take NOE intensities resulting from spin diffusion into account, only backbone NOE contacts are roughly scaled according strong, medium, and weak classification.

- Backbone torsion restraints from TALOS predictions As sequential backbone assignment is available, prediction for backbone ϕ,ψ torsions can readily obtained in an automatic manner using the TALOS program [25].

- Hydrogen bonds can be inferred from backbone chemical shift assignment in interrelation with characteristic H^N-H^N NOE patterns as well as amide proton attenuation factors (if available). This step can be easily automated.

The feasibility of the protocol was tested using a set of various proteins.

2.7 A direct method protocol employing residue type specific isotope labeling for rapid structure validation and model-driven sequential assignment

In contrast to conventional NMR structure analysis, where various triple resonance experiments need to be recorded in order to obtain the sequence-specific assignment necessary to interpret the NOESY spectra, 'direct' methods [89], [50], [5] initial generate a structure of detected but unassigned (and unconnected) atoms, which serves as template for subsequent automated resonance assignment. These methods rely on a very large number of unambiguous NOE distances, which have to be determined with high accuracy (e.g. from relaxation matrix refinement) to obtain a 3D structure of adequate resolution.

While the (indirect) standard procedure is almost guaranteed to work (provided sufficient protein solubility and stability), direct methods can potentially accelerate the task of rapid fold evaluation (as demanded in structural genomic projects, e.g. to verify a proposed homology model),

where sequential NMR assignment is obtained as a by-product in the course of the structure calculation rather than being its prerequisite.

As mentioned above, although such 'direct' approaches sound very appealing *per-se*, a high number of accurate NOE contacts have to be unambiguously defined to obtain at least a low resolution image of the correct protein fold.

Therefore a hybrid approach was developed and tested in this work, which requires a minimum number of experimental NOEs only. In turn it can then be used in combination with a minimal amount of triple-resonance data to support and drive the sequential assignment process and the NOE identification by making reference to that low-resolution structure model. Once the sequential assignment has been established, the structure determination proceeds in a conventional way using standard covalent protein templates. In that case the experimentally NOEs previously assigned can be sufficient to obtain the protein backbone fold and the model obtained such can serve as a starting point for further refinement by iterative assignment of ambiguous NOEs (ARIA) [128] or similar algorithms (see chapter 1.5).

The (unassigned) experimentally NOE database used in that procedure encompasses beside backbone amide $^{1}H^{N}$-$^{1}H^{N}$ contacts, which can be readily unambiguously obtained from 4D ^{15}N-^{15}N edited NOESY-HSQC experiment, entries of methyl group bearing and (optionally) aromatic residues, which preferentially located in the protein core or at protein-protein interfaces, where they give rise for valuable NOEs and therefore greatly improve the precision of the structural model [78]. In that context residue and position selective, isotope labeling techniques are particular convenient because they allow for spectral simplification and editing of multidimensional NOE spectra at the same time. Labeling patterns were designed to facilitate rapid identification of valine, isoleucine and sequentially neighbored residues through residue type edited triple-resonance experiments [182].

Using the principles described in section 2.5.3, the identification of methyl bearing residues (as isoleucines, leucines, and valines) and the unambiguous linking to their intra-residues backbone amide $^{1}H^{N}$ resonance is also possible through careful analysis of their $^{13}CH_3$ NOE patterns, which must be complementary to the corresponding $^{1}H^{N}$ NOE patterns. Residue types can be derived from $^{13}C^{\alpha}$ and $^{13}CH_3$ chemical shifts.

As mentioned in section 2.5.1, normally the $^{13}C^{\alpha}$ connectivity information from 3D HNCA [75] (in combination with 3D HN(CO)CA) [53] data alone is not sufficient for unambiguous sequential assignment due to the large shift degeneracy of the $^{13}C^{\alpha}$ backbone shifts. Further information from other spin types is usually required to make the sequential connectivity unambigu-

ous ($^{13}C^{\beta}$ chemical shifts from 3D HNCACB/3D CBCA(CO)NH [195] or $^{13}C^{\alpha}$ chemical shifts employing 3D HNCO/3D HN(CA)CO [75]). However, the performance of these experiments may prove more difficult especially in larger molecular weight systems due to fast and efficient transverse ^{15}N and ^{13}C transverse T_2 relaxation. While HNCA (and HNCO) and potentially HN(CO)CA spectra can still be acquired with reasonable sensitivity in the absence of sidechain deuteration even in large molecular weight systems, the more complex experiments require full or partial deuteration especially in conjunction with TROSY methodology [138], otherwise the quality of the other triple-resonance datasets is seriously compromised by fast relaxation. In contrast, even in that cases NOESY spectra of good quality are normally obtained without large experimental difficulty. Then, NOE connectivities can principally provide the missing information required for unambiguous sequential linkage. In the absence of a structural model only sequential to mid-range NOEs are interpretable. The information content of a NOE data set is further enhanced if a three dimensional protein model is available, in whose context all $^{1}H^{N}$-$^{1}H^{N}$ NOEs, including long-range and characteristic tertiary contacts (e.g. across β-strands), can be used for sequential linkage and placement of spin systems onto a protein structure. This process is greatly improved when a-priori residue typing for a number of spin-systems (e.g. by selective labeling) is possible, which will subsequently serve as pivotal points for sequential assignment and as check points for validation (see section 2.5.1).

The presented (direct) method here facilitates the automated generation of a low- to mid resolution protein backbone folds for globular proteins requiring only 3D HNCA/3D HN(CO)CA input data in conjunction with NOE derived unassigned atom density clouds using different chemical isotope labeling strategies. The approach is to some extend error-tolerant with respect to data incompleteness, limited precision of the peak picking, and structural errors caused by misassignment of NOEs. A flowchart of the procedure is given in figure 2.11.

2.7.1 Input Data and Systems Studied

A set of seven proteins with different fold topology, extent of accessible chemical shifts, backbone chemical shift dispersion, and molecular sizes served for testing the procedure. Almost complete chemical shift assignments and high-resolution NMR structures of Cyclophilin D [158], ICln [159] and LIM1(CRP2) (data not published, see see chapter 4) were originally manually obtained in the course of conventional (indirect) structure determination projects. The experimentally derived NMR distance restraints used here were exclusively generated from

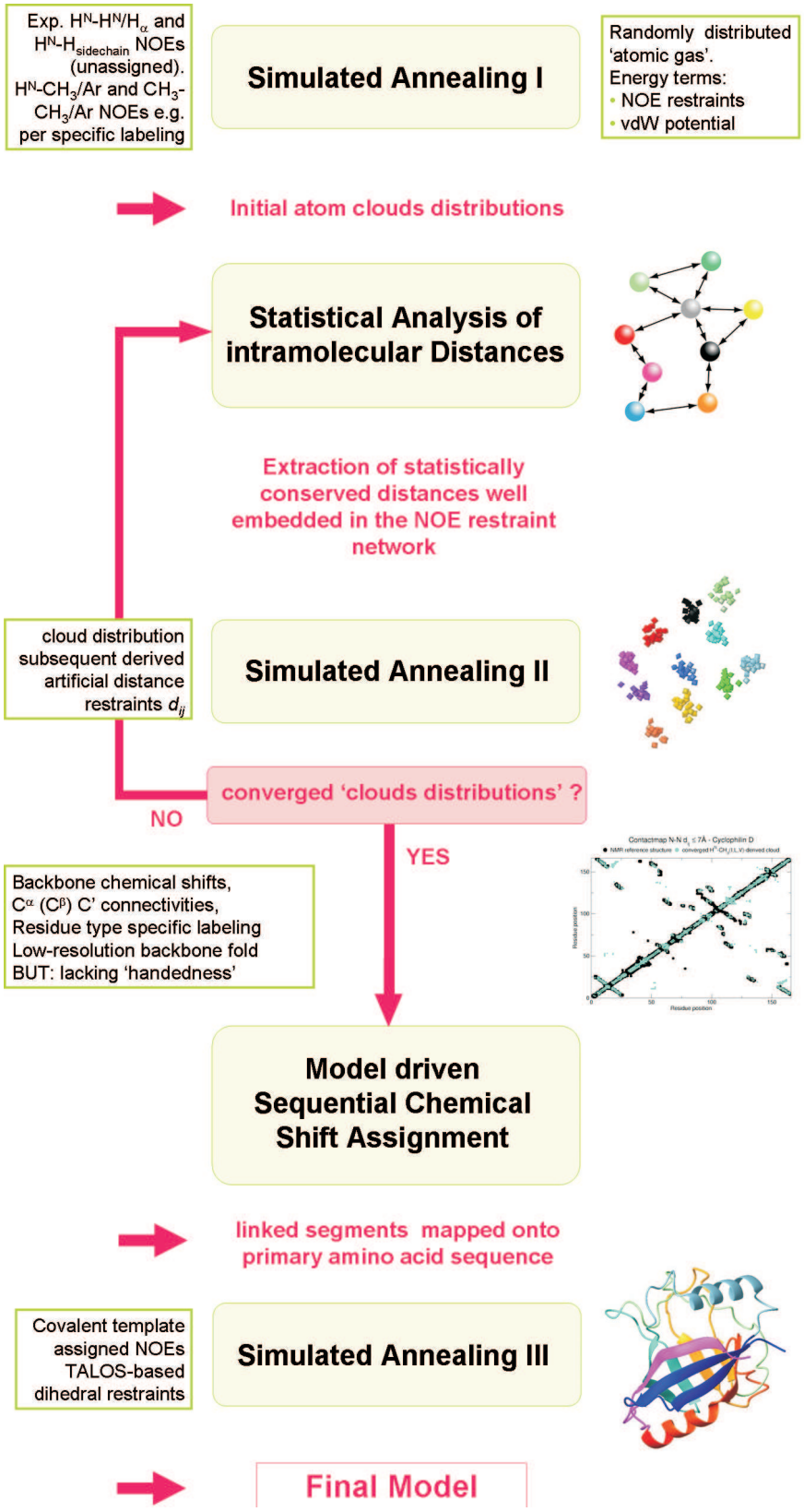


Figure 2.11: Summary of the direct method protocol developed in this work.

unambiguously assigned NOESY cross peaklists available from those conventional structure determinations.

Experimental NOE databases used in this procedure contained only NOEs, which can be obtained with good sensitivity and resolution in 3D and 4D heteronuclear NOESY experiments using either globally doubly (^{15}N/^{13}C) labeled or position specifically labeled (using suitable precursor molecules) proteins. They contained backbone H^N-H^N and (optional) H^N-H_α NOE distances, backbone to sidechain H^N-$H_{SideChain}$ (were $H_{SideChain}$ denotes methyl CH_3(ALA,THR,ILE,LEU,VAL), aromatic H_δ/H_ϵ/H_ζ(PHE,TYR), and sidechain NH_2(ASN,GLN) atoms, respectively), and sidechain to sidechain CH_3-CH_3(ALA,THR,ILE,LEU,VAL) distances, and CH_3(ALA,THR,ILE,LE NH_2(ASN,GLN) contacts.

Using Cyclophilin D as an example the availability of a even more sparse set of experimentally NOE distances involving solely $^1H^N$ and CH_3 groups (case a), additionally supplemented with distance restraints involving aromatic H_δ/H_ϵ/H_ζ(PHE,TYR) protons (case b) was simulated and their influence on the performance of the procedure studied.

To investigate the influence of further different protein fold topologies idealized NOE data sets were simulated for a number of different proteins of varying secondary and tertiary structures. Backbone chemical shifts and structural data for Calmodulin, Carbonic Anhydrase I, Plastocyanin, and Thioredoxin were taken from BioMagResBank and Protein Data Bank, respectively. For those test set members theoretical NOE distances comprising H^N and CH_3(ILE,LEU,VAL) atoms only were inferred directly using an interatomic distance cutoff of 6Å, within NOE cross peaks were assumed to occur.

2.7.2 Restraint Classification

Experimental backbone NOEs, encompassing H^N and (optional) H_α protons were roughly classified according their NOESY cross peak volumes as strong, medium and weak, corresponding to distance limits of 1.8-2.8Å, 1.8-3.4Å and 1.8-5.0Å, respectively. Taking inconvenience of quantification NOEs belonging to sidechain atoms into account, distances related to backbone to sidechain and inter-sidechain NOEs (H^N-CH_3, CH_3-CH_3) were uniformly restrained between 1.8-6.0Å.

For weak NOEs small upper limit violations of up to 2.0Å i.e. $<$ 7.0Å for backbone NOEs and $<$ 8.0Å for sidechain NOEs are only penalized with a very small force constant. This applies only a weak bias towards the distance range $<$ 5Å (for backbone-backbone) / 6Å (NOEs of sidechains)

while practically allowing distances up to 7-8Å accounting for possible effects of spin-diffusion. In practice this is implemented by using a sum of two distance restraints, 1.8-7.0/8.0Å(subset A) and 1.8-5.0/6.0Å (subset B, related to the same NOE distance) with two different force constants. This was empirically found to improve the quality of the derived structures.
For intra-residue NOE contacts (e.g. between prochiral $CH_3^{\gamma 1,2}$ of valine residues and $CH_3^{\delta 1,2}$ of leucine residues) as can be identified through high intensity cross peaks in the NOESY spectra (or from the analysis of HCCH-COSY/TOCSY experiments, if available), an upper bound of the two related distance restraints were set to 3.0 and 5.0Å, respectively. The protons of the methyl groups -CH_3, glycine $H^{\alpha 1}$/$H^{\alpha 2}$, aromatic H_δ/H_ϵ/H_ζ(PHE,TYR), as well as sidechain amides -NH_2 (ASN, GLN) were represented as grouped pseudo-atoms H_* in the relevant NOE databases and molecular structure templates, respectively. Analogously, aromatic residues can be identified by their characteristic intra-residue NOE contacts or by HCCH-COSY/TOCSY experiments, if available. Corresponding distance restraints were defined between aromatic intra-residue Phe ($H_{\delta *}$ to $H_{\epsilon *}$, $H_{\delta *}$ to $H_{\zeta *}$,and $H_{\epsilon *}$ to $H_{\zeta *}$) or Tyr ($H_{\delta *}$ to $H_{\epsilon *}$) pseudo-atoms in order to achieve consistent spatial vicinity and a roughly linear arrangement of these entities. $^1H^N$, ^{15}N, $^{13}C^\alpha$, $^{13}C^\beta$,^{2}D, β-sheets, α-helices, $^{13}CH_3$,aromatic H_δ/H_ϵ/H_ζ(PHE,TYR)

2.7.3 Structure Calculations: Simulated Annealing I

An ensemble of spatial atomic distributions calculations were performed (round 0) using NIH-xplor software version 2.9.6 or newer [164]. A simulated annealing protocol was applied to a gaseous randomized distribution of atoms connected only by distance restraints, and which were initially randomly spread out over a cube of 50 to 100Å. At this point, gross errors in the NOE list can be detected by analyzing the distributions of NOEs connectivities/pathways between all the atoms of the distance restraint database. Only a soft square NOE potential and a hard sphere van der Waal energy term but no covalent (bond, angle, torsion, improper) energy terms were applied. Only atomic clouds with zero restraint violations were selected for further analysis.

2.7.4 Statistical Analysis of Conserved Distances

When analyzing the distribution of internuclear distances of the atomic distribution generated above, internuclear distances between two atoms i and j, that are not connected by a direct NOE restraint, are nevertheless well-conserved. This happens when both atoms concerned

are connected by a short unbroken chain of individual NOEs, whose related atoms together with atoms i and j appear to be well embedded in the network of NOEs. The number of intervening NOEs is typically referred to as the NOE path length. Such consistently conserved distances can potentially serve as restraints for further rounds of structure calculations. But being conserved is necessary but not a sufficient criteria for derivation of a new distance restraint compatible with high resolution structure priory obtained in a conventional manner. Missing NOEs and the roughly classification of their distance restraints leads to more expanded structures.

Figure 2.12.A, figure 2.12.C visualize the tendency of systematically overestimation of such distances. Internuclear distances (black) extracted from a family of atomic distributions ('clouds') with their target values measured in a reference structure. Because of this a filter was applied selecting only those conserved distances (cyan) well embedded in the input NOE network (see figure 2.12.B, figure 2.12.D).

Conserved distances were only converted into distance restraints D_{ij}, when the two involved atoms i and j were connected by a NOE path length l of 2 $\leqq$ l $\leqq$ 6 individual NOEs and when they were conserved within relatively narrow margins. Only triatomic distances r_{ij} with a standard deviation σ_{ij} $\leqq$ 20% of their mean values d_{ij} (calculated over the whole family of atomic clouds) or $\leqq$ 5Å in absolute numbers were considered. The corresponding upper bound d_{upper} and lower bound d_{lower} were set to $\bar{d}_{ij}$ + 2σ_{ij} unless σ_{ij} became smaller than 10% of the mean distance $\bar{d}_{ij}$. In that case d_{upper} and d_{lower} were redefined as $d_{ij}\pm$ 0.2d_{ij}, in order not to artificially 'overrestrain' the system.

Furthermore, for the rarely occurring cases when the shortest individual distance in the family of calculated structures $\min(r_{ij})$ was smaller than d_{lower} or the largest individual distance of the family of clouds and $\max(r_{ij})$ was longer than d_{upper} the corresponding *artificial* distance restraint A_{ij} was redefined to range from $[\min(r_{ij}) - \sigma_{ij}]$ to $[\max(r_{ij}) + \sigma_{ij}]$

2.7.5 Simulated Annealing II and Statistical Analysis II

The artificial distance restraints A^1_{ij} derived from the initial family of atom clouds, as described above (iteration round 0), were merged with the original NOE derived distance restraints and used for a subsequent round of structure calculation (iteration round 1) using essentially the same protocol as used for calculation of the initial structure ensemble. However, the applied force constant $f_{artificial}$ of those artificial statistical A^0_{ij} was scaled down by a factor of five rela-

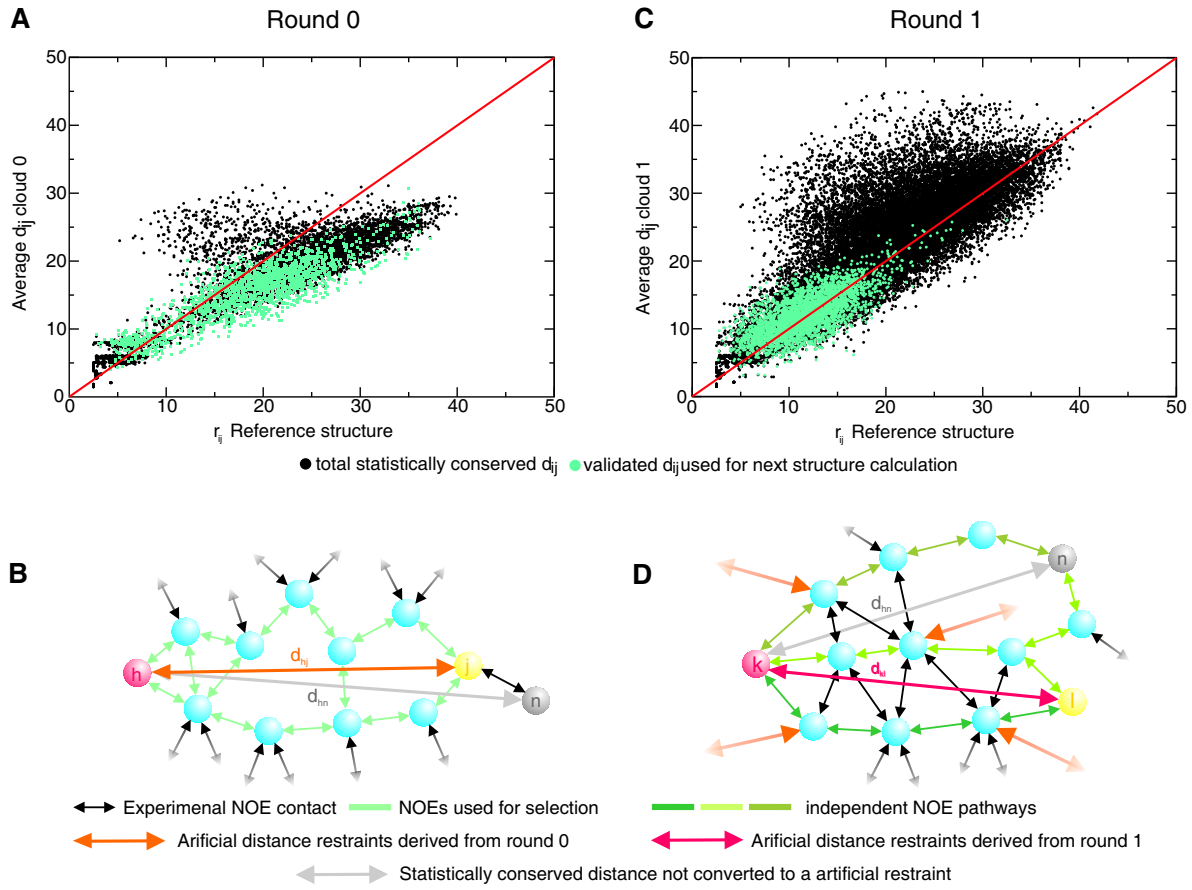


Figure 2.12: Distance plot for Cyclophilin D. (A) Atomic distribution ('cloud') distances r_{ij} averaged over an ensemble of 50 structures generated from a random gaseous proton distribution using the (experimental) input NOE distance restraints only (Simulated Annealing I, round 0) and found to be statistically conserved are plotted against those of the reference structure (black). Without the presence of restraints subset B (implemented to bias a distance toward more statistically signifi cant values, see 'Restraint Classifi cation') most of the derived distance correlation in subsequent rounds would tent to be considerable overestimated. Applying more stringent selection criteria additionally depending on the experimental input NOE density (Statistical Analysis I) lead to a improved set of distance restraints (cyan) where the majority of erroneous overestimated d_{ij} are effi ciently fi ltered out and hard out liners are completely removed. (B) Illustration of the selection criteria used in the Statistical Analysis I. Individual experimental input NOEs are represented by black double arrow, whereas statistically conserved distances selected and converted to distance restraints A_{ij} for subsequent structure calculation and those rejected because not fulfi lling the selection criteria are colored in red and gray, respectively. (C) Atomic 'cloud' distances r_{ij} averaged over an ensemble of 50 structures generated from a random gaseous proton distribution (Simulated Annealing II, round 1) which was calculated in presence of both experimental NOE distance restraints and those extracted as conserved in the previous round and converted to A_{ij}s are plotted against the related distances found in the reference NMR structure (black). Due to the presence of A_{ij} for deriving this structure data only those statistically conserved distances from round 1 are selected (colored cyan) well embedded in the experimental NOE network (D).

tive to the force constant f_{NOE} of the direct experimentally derived NOE contacts to ensure that all these original experimental NOE distance restraints remain fulfilled. The resulting ensemble of atom cloud distributions showed no violation $>$0.5Å. From the best structures (with respect to lowest NOE energy values) new statistically conserved distances r_{ij} were extracted and selected to yield another new set of artificial distance restraints A^2_{ij}. Because of the presence of the A^1_{ij} during the previous, first round of structure calculation more stringent criteria had to be applied in their selection:

- The two distance related atoms i and j should be connected by a NOE path length of l of $2 \leqq l \leqq 4$ individual NOEs.
- At least two independent NOE pathways should be present (i.e. pathways that do not share any atom along their individual NOE route between the atoms i and j with exception of their direct NOE partners).
- All atoms along the NOE pathway had at least two additional NOE contacts with surrounding atoms.
- Finally, all distances $d_{i,j\rightarrow k}$ between atoms i and j to any third atom k of the NOE pathway were larger than the lower bound of the experimental NOEs.

Applying these newly-derived artificial distances restraints A^2_{ij} in another round of structure calculations a new set of atomic clouds was calculated (iteration round 2). The procedure of structure calculations and identification/selection of conserved distances can be repeated and iterated until convergence, i.e. an (almost) constant number $N_{conserved}$ of statistically conserved distances for a distinct quantum of atoms (predominantly located in secondary elements of high NOE density) was observed and no significant changes in $N_{conserved}$ of other atoms was found. Typically, convergence occurred after two iterations.

Applying the structure generation protocol outlined above convergent atomic densities were be obtained, which resembled a low-resolution images of a protein or parts thereof. Converged parts of a protein structure, which are superponable, could be identified by analyzing the matrix of mutual distances between atoms r_{ij}. Atoms connected by distances r_{ij} offering small variations of σ_{ij} values indicate regions with defined structure.

Similar to Distance Geometry [60], [24] the resulting atomic clouds clustered in two degenerate families of structures, which are related to each other as mirror images of each other, because the NOE term U_{NOE} is invariant to inversion or reflection. Given the typical resolution of the

atomic clouds the correct protein fold could not be distinguished from its inverse one and the two families of structures remained degenerate. The three dimensional proton clouds obtained were of great value in the process of sequential signal assignment as they were capable of improving the information content of triple resonance data, as only distance information is used, both the families of structures can be used equally well.

2.7.6 Model Driven Sequential Chemical Shift Assignment

To demonstrate the usefulness of low-resolution atom density clouds and asses their impact in the context of sequential assignment they were included together with backbone chemical shift data into an automated Monte Carlo simulated annealing (MCSA) routine for sequential chemical shift assignment. The program MONTE [63](see section 2.5.1) was applied for that purpose.

The input data comprised ^{15}N edited 3D (and 4D) NOE information and backbone $^{13}C_i^{\alpha}$ and $^{13}C_{i-1}^{\alpha}$ shifts (and for comparison $^{13}C_i^{\beta}$, $^{13}C_{i-1}^{\beta}$, $^{13}C_i^{'}$ and $^{13}C_{i-1}^{'}$ chemical shifts when available) classified by $^{1}H^{N}$-^{15}N spin system. Furthermore, either the primary sequence with or without a three-dimensional PDB model or a atomic density structure was supplied. Residue type specific labeling (when available) was included as additional boundary condition.

A number of different input scenarios were tested for comparison. Different extends of available ^{13}C backbone chemical shifts ($^{13}C^{\alpha}$/$^{13}C^{\beta}$ or $^{13}C^{\alpha}$ alone), different extends of precision and completeness of ^{13}C backbone chemical shifts were used. The impact of inclusion of a three dimensional model was evaluated by inclusion or omission of atomic density structures of different resolutions. In a similar way different residue specific labeling strategies were evaluated, too. 'Reliability' of the obtained assignments was derived from their reproducibility in a number of independent simulated annealing Monte Carlo cycles. The initially manually performed chemical shift assignments serves as guideline for the correctness of the obtained SA-MC results.

2.7.7 Simulated annealing III

With the sequential protein backbone available the structure calculation was repeated using a covalent protein template and assigned NOE distance restraints. The connection of sidechain methyl groups to their backbone amide functions could be established applying the automated assignment approach described in section 2.5.3. Only NOEs originating from as-

signed residues from the parts of the protein, that were deemed to be uniquely and reliably assigned, were included. The remainder of the protein structure was restrained solely by empirical database potentials and restrictions imposed by the covalent bond geometry.

2.8 Automated Protein-Ligand Binding Site Mapping Using Unassigned Triple-Resonance Data

The identification of protein ligand binding site is of out most interest in rational drug design. Conventional NMR approaches rely on sequentially-assigned NMR data and the analysis of differential residue-specific chemical shift changes upon ligand binding. In this work a algorithm for NMR-based protein-ligand binding site mapping using *unassigned* triple-resonance data developed by Robert Konrat (paper in preparation) was tested using a selected set of several already deposited protein-ligand systems.

The program is part of a software platform, which contain beside tools for triple-resonance data analysis and Monte-Carlo (MC) type sequential signal assignment [65] additionally a module for NOE-based assignment of methyl groups as described in section 2.5.3.

Identification of protein ligand interaction sites via NMR, a suitable technique for throughput screening, is very valuable for finding lead compound and the establishing of common binding modes. Traditional chemical shift based approaches rely on uniform ^{13}C,^{15}N labeling of one components (and their complete sequential resonance assignment) while leaving the second unlabeled [15]. The ligand binding site is inferred from backbone and/or sidechain chemical shift changes between the free and ligand bounded form, whereas preferentially carbon resonances instead of backbone ^{1}H,^{15}N resonances are employed for detection because of increased chemical shift dispersion.

In the conceptually different approach (termed ROSA for RObust binding Site Allocation) tested here, the ligand binding site is identified information and assuming a local ligand binding site (i.e. the absence of allosteric effects)by using*unassigned* NMR triple-resonance data and a 3D structure model of the protein. Spin system information for a residue i ($^{1}\mathrm{H}_i^N$, $^{15}\mathrm{N}_i$, $^{13}\mathrm{C}_i^{\alpha}$ and $^{13}\mathrm{C}_{i-1}^{\alpha}$ chemical shifts supplemented by $^{13}\mathrm{C}_i^{\beta}$, $^{13}\mathrm{C}_{i-1}^{\beta}$, and $^{13}\mathrm{C}_i^{'}$, and $^{13}\mathrm{C}_{i-1}^{'}$resonances, if available) is derived from readily obtainable 3D HNCA/3D HN(CO)CA spectra and additionally from 3D HNCO/3D HN(CA)CO and/or HNCACB/3D CBCA(CO)NH experiments, if available. Optionally, sequential $^{1}\mathrm{H}_i^N$-$^{1}\mathrm{H}_i^N$ NOE contacts can be used. If spectral resolution allows, the identification of spin system that exhibit chemical shift changes upon ligand binding is per-

formed by a 2D $^{1}H_{i}^{N}$, $^{15}N_{i}$ HSQC experiment(s).

All possibilities for sequential connections of these spinsystems S by chemical shift matching (or additionally, by sequential $^{1}H_{i}^{N}$-$^{1}H_{i}^{N}$ NOE contact analysis) are evaluated. For mapping on the primary sequence a fragment size of four sequentially connected spinsystems (S_{i-2}-S_{i-1}-S_{i}-S_{i+1}) and average residue specific ^{13}C chemical shift values taken from the BMRB database (http://www.bmrb.wisc.edu) are used.

To achieve an unambiguous placement of these residue quadruplets Q onto the primary sequence distance information from the 3D structure model is exploited (see figure 2.13). Assuming that all related residues with distinct chemical shift changes upon binding are in close spatial neighborhood to the binding site (and therefor to each other) ROSA evaluates the structural vicinity for all putative backbone positions (of the shifted spinsystem mapped as frequency quadruplet) using a user-defined C^{α}-C^{α} distance cutoff.

A subsequent calculated binding site identification score depends on the number of jointly shared neighborhood residue positions assigned to shifted spin systems. These residue localizations on the 3D structure are highlighted and considered to be part/or next the ligand binding site (assuming the absence of allosteric effects).

The performance of this routine was assessed using a series of test complexes, whose *apo* and *holo* chemical shifts as well as a related structure model were existent at the BMRB (BioMagResBank, www.bmrb.wisc.edu) and PDB (protein data bank, http://www.pdb.org), respectively.

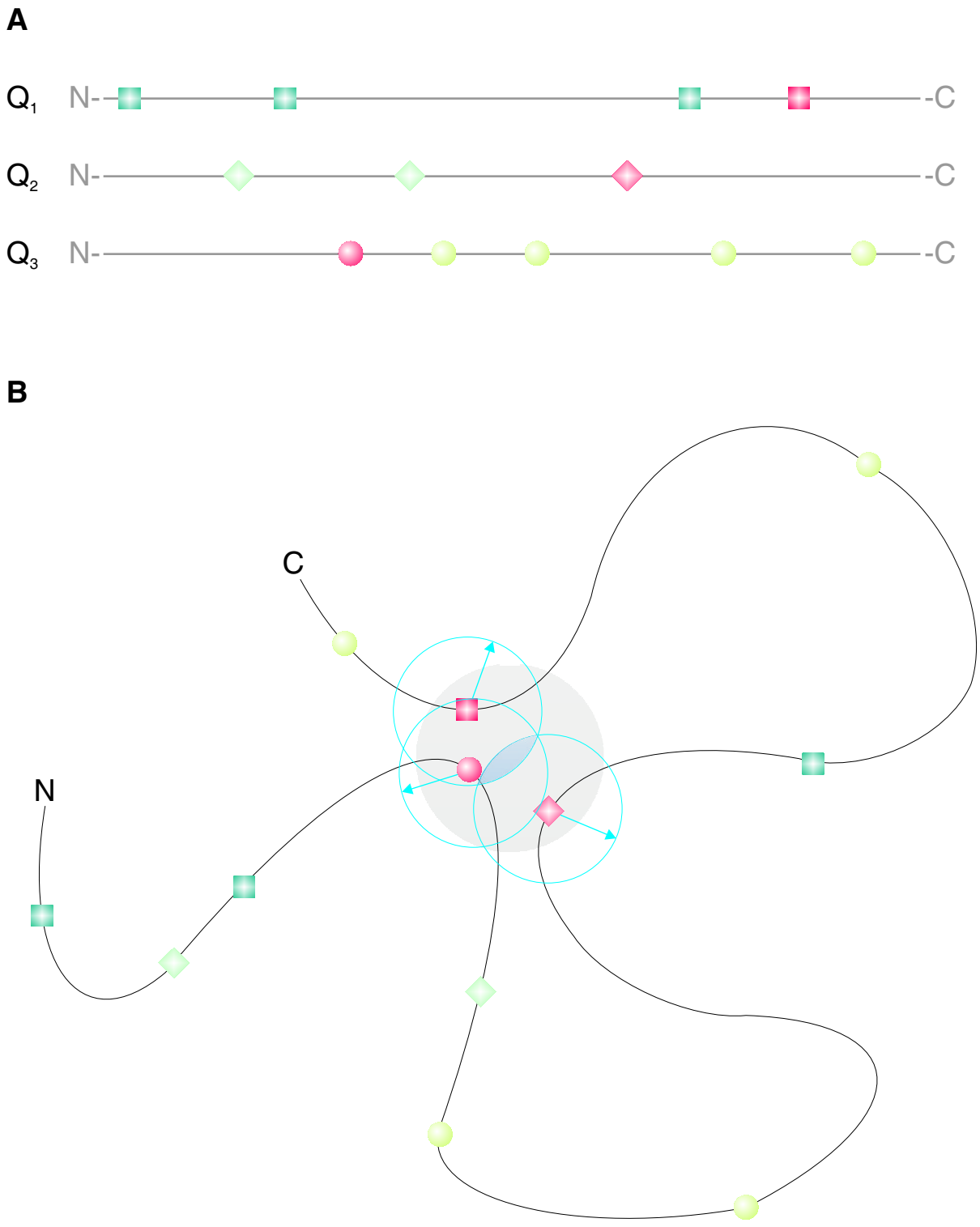


Figure 2.13: Scheme for automated binding site prediction tool (ROSA). (A) Primary sequence mapping possibilities of spinsystem quadruplets Q_1-Q_3 (e.g. connected by their containing $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ and/or $^{13}C^{'}$ resonances) comprising residues whose $^{1}H^{N}$ and ^{15}N chemical changes upon ligand binding. (B) Correct Q_x spinfragment positions at the binding site (gray area) and their alternative assignments are drawn as red and green tinted symbols, respectively. For further restraining of the potential assignments a structural neighborhood criteria is applied demanding for a common slice volumina (within a user-defi ned distance cutoff) of the mapped residue positions.

Chapter 3

Experimental

3.1 Samples and NMR parameters

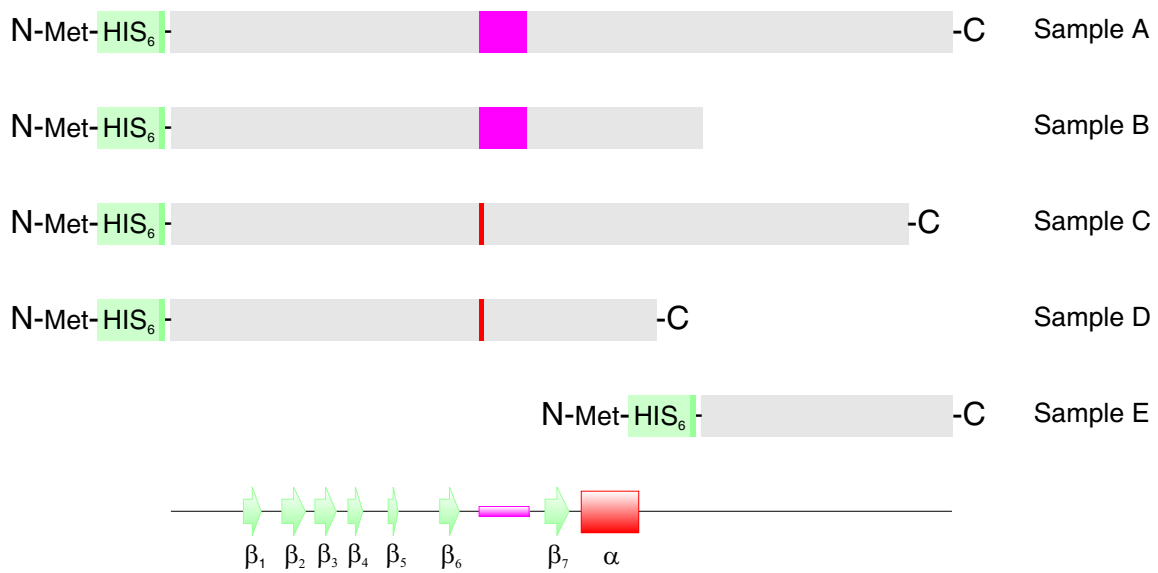
3.1.1 ICln

A broad variety of samples constructs for structural and dynamical characterization of ICln were provided by the Paulmichl group. Some of these were applied in ligand interaction studies. Where necessary, ultra centrifugation through Amicon ultra 10 MW filters was performed until a concentration of $\sim$ 1mM suitable for NMR experiments was achieved. Except for some ligand titration experiments 250 μl Shigemi tubes were used for NMR data acquisition. All buffers contained 10% D_2O (v/v) for shim locking.

3.1.1.1 Samples for structure determination

Beside a preparation of the 235 residue wild type ICln (sample A) several mutants were available for NMR investigations. For purification purposes each sample contained beside a starting methionine a N-terminal 6xHis tag linked by a two residue spacer (Leu-Glu) to the ICln amino acid sequence. The NMR samples for sequential assignment and structure determination were uniformly ^{13}C,^{15}N labeled by adding $^{15}NH_4Cl$ and [^{13}C]-*D*-Glucose to the minimal media.
From CD-spectra and primary sequence analysis the N-terminal part was expected to be preponderant folded. This was found in contrast to the C-terminal part of ICln predicted to exhibit random coil structure and whose stability in solution was found to be poor. Thus, initial a truncated construct encompassing the N-terminal part residue Met1 to residue Gln159 was investigated (sample B). In another version of this mutant (sample C) a 18 residue stretch (residue Glu89 to residue Ile107) subsequent found to be highly dynamical and disordered was

A



B

```
MSFLKSFPPP GSAEGLRQQQ PETEAVLNGK GLGTGTLYIA ESRLSWLDGS
GLGFSLEYPT ISLHAVSRDL NAYPREHLYV MVNAKFGEES KESVAEEEDS
DDDVEPIAEF RFVPSDKSAL EAMFTAMCEC QALHPDPEDE DSDDYDGEEY
DVEAHEQGQG DIPTFYTYEE GLSHLTAEGQ ATLERLEGML SQSVSSQYNM
AGVRTEDSTR DYEDGMEVDT TPTVAGQFED ADVDH
```

Figure 3.1: Samples of ICln used for NMR analysis. (A) Representation of the constructs. In some cases the high dynamical and structural independent region (containing loop β_6/β_7, colored in magenta) was replaced by a two residue linker (red) for spectra simplifi cation purposes. At the bottom the secondary structure elements are shown as derived from secondary chemical shift and NOE pattern analysis. (B) Complete wild type primary sequence of *Canis familiaris* (dog) ICln.

replaced by a two residue linker (Gly-Ser) for simplifying the spectra and to detect any contingent structural consequences when missing that fragment. Furthermore, an analog construct that additionally contained the complete C-terminus of wild type ICln was analyzed (sample D). Possible preferences for a secondary structure of the C-terminal part were analyzed by a sample comprising residue Gln159 to residue His235 (sample E). All samples for resonance assignment and structure determination contained 150 mM NaCl in 25 mM K_2HPO_4, at pH 7.0 (puffer A). An overview of the sample constructs is given in figure 3.1.

3.1.1.2 Samples for phosphorylation state and ligand interaction studies

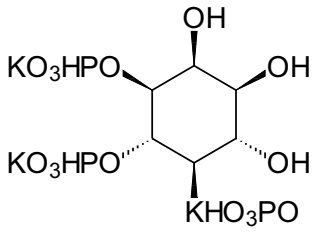
For titration experiments both the protein and the ligand were solved in the identical buffer.

Phosphorylation state analysis

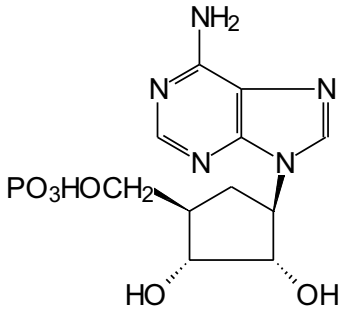
Since ICln offers a number of consensus sites for phosphorylation by protein kinase C (beside protein kinase A and G; casein kinase I and II), the phosphorylation state of an uniformly ^{15}N-labeled truncated form of wild type ICln containing residue 1 to residue 158 (sample B1) after *in vitro* PKC treatment was analyzed.

*^{15}N-labeled E98-D112-GS-Mutant (sample C) with D-*myo *Inositol-3,4,5-trisphosphate (D-*myo *Ins(3,4,5)P$_3$)*

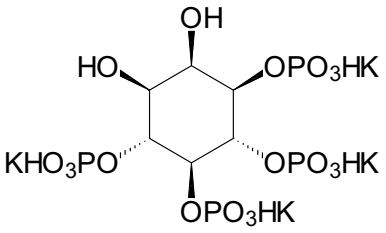
A solution of 100 μg D-myo-Inositol 3,4,5-trisphosphate.tripotassium salt (M_w=534.4; $C_6H_{12}O_{15}P_3.K_3$) in 150 μl buffer (150 mM NaCl, 25 mM K_2HPO_4, pH 7.0) was added to 100 μl $\sim$ 1mM protein. The resulting molar protein to ligand ratio was 0.5:1.

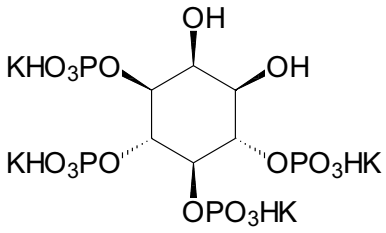


D-*myo*-Inositol
3,4,5-tris-phosphate-tripotassium



Adenosine-5'-monophosphate





DL-*myo*-Inositol 1,4,5,6-tetrakis-phosphate.tetrapotassium

Figure 3.2: Structures of potential ICln ligands applied in titration experiments.

*^{13}C,^{15}N-labeled truncated wild type ICln (sample B) with DL-*myo *Inositol-1,4,5,6-tetrakis-phosphate (DL-*myo *Ins(1,3,4,5)P$_4$)*

100 μg DL-*myo* Inositol-1,4,5,6-tetrakisphosphate.tetrapotassium salt (M_w= 652.1; $C_6H_{12}O_{18}$

$P_4.K_4$;) were added as solid to 250 μl protein solution in 150 mM NaCl, and 25 mM K_2HPO_4 at pH 7.0, thus obtaining a molar protein to ligand ratio of 1:0.57 ligand.

^{15}N-labeled E98-D112-GS-Mutant (sample C1) with 5'AMP

A sample of 600μl 0.4mM ^{15}N-labeled E98-D112-GS-Mutant in 150 mM NaCl, and 25 mM K_2HPO_4 at pH 7.0, was titrated with a 38,9 mM solution of 5'AMP (M_w=347.22; $C_{10}H_{14}\ N_5O_7P$).

^{15}N-labeled E98-D112-GS-Mutant (sample C1) with Ca^{2+}

A sample of 600μl 0.4mM ^{15}N-labeled E98-D112-GS-Mutant in 25 mM MOPS (3-(N-morpholino)-propanesulfonic acid; $C_7H_{15}NO_4S$) at pH 7.0 and 150 mM NaCl was titrated with a 250 mM solution of $CaCl_2$ ($CaCl_2$ $2H_2O$; M_w=147.01).

^{15}N-labeled truncated wild type ICln (sample B) with CHAPS

A protein solution of $\sim$1mM ^{15}N-labeled truncated wild type ICln in 25 mM MOPS at pH 7.0 and 150 mM NaCl was titrated with 20% solution of CHAPS (3[(3-Cholamindo-propyl)-dimethylammonio]-1-propanesulfonate; M_w=614.9; $C_{32}H_{58}N_2O_7S$).

^{15}N-labeled truncated wild type ICln (sample B) with $\alpha_2\beta$ Integrin peptide

A sample of $\sim$2mM ^{15}N-labeled truncated wild type ICln in 150 mM NaCl and 25mM MOPS at pH=7.0 was mixed with a $\approx$? mM solution of a 11 residue synthetic peptide encompassing the membrane-proximal regions of the platelet $\alpha_2\beta$ integrin protein (M987-WKVGFFKRN-R997; M_w=1468.7; $C_{69}H_{105}N_{21}O_{13}$; pI=12.02). Although more soft ion strength condition were applied by using MOPS instead of phosphate as puffer system, some precipitation was observed in the integrin peptide solution prior mixing. The supernatant was taken for titration.

Titration experiments with LSm4

First, 250μl $\sim$2mM ^{15}N-labeled truncated wild type ICln (sample B) in 150 mM NaCl, and 25 mM K_2HPO_4 at pH 7.0 was warily mixed with 250μl $\sim$2mM LSm4 protein solution, resulting in some protein precipitation. The supernatant was used for data acquisition.

Then, a sample of 250 μl $\sim$2mM ^{15}N-labeled E98-D112-GS-Mutant (sample C1) protein (lacking the first acidic domain in the $\beta_6\beta_7$ loop region) in the same buffer was gently mixed with 250 μl LSm4 protein solution of the same concentration. No precipitation was observed.

Finally, this experiment was repeated using a solution of 125 μl $\sim$2mM wild type ICln for titration with 125 μl $\sim$2mM under identical buffer condition.

^{13}C,^{15}N-labeled C-terminal ICln with the osmolyt L-argenine

A sample of $\sim$1mM ^{13}C,^{15}N-labeled C-terminal ICln in 150 mM NaCl and 25 mM K_2HPO_4 at pH 7.0 was mixed with a $\sim$ 2mM solution of L-argenine (L-argenine.hydrochloride; M_w=174.2) for measurement at 281°K and 298°K.

3.1.2 Cyclophilin D

A sample of uniformly ^{13}C,^{15}N-labeled wild type Cyclophilin D (165 residues, no tag) was expressed and provided by Simon Ruedisser of the Novartis Schwitzerland Company[1], which contained $\sim$ 800μM protein, 50mM Na_2HPO_4 buffer at pH 7.0 and 10%D_2O. This preparation was empoyed for all NMR experiments using a 250 μl Shigemi tube for NMR data acquisition. The sequence is given in figure 3.3.

```
MQNPLVYLDV GADGQPLGRV VLELKADVVP KTAENFRALC TGEKGFGYKG
STFHRVIPAF MCQAGDFTNH NGTGGKSIYG SRFPDENFTL KHVGPGVLSM
ANAGPNTNGS QFFICTIKTD WLDGKHVVFG HVKEGMDVVK KIESFGSKSG
KTSKKIVITD CGQLS
```

Figure 3.3: Primary sequence of *homo sapiens* (human) cyclophilin D.

3.1.3 LIM1 of CRP2

A recombinant uniformly ^{13}C,^{15}N-labeled LIM1 domain poly-peptide comprising the N-terminal part (residue 1 to residue 81) of wild type quail CRP2 was prepared at the Klaus Bister group[2]. The sample for structure characterization by NMR included 2.4 mM protein, 20 mM K_2HPO_4 at pH 7.2, 50 mM KCl, 0.5 mM dithiothreitol, and 10%D_2O. That sample was used for all NMR experiments using a 250 μl Shigemi tube for NMR data acquisition.. The sequence of the investigated peptide including the two tetrahedral Zn^{2+}-coordinating sites of the type Cys-Cys-His-Cys and Cys-Cys-Cys-Cys, respectively) is shown in figure 3.4.

3.1.4 BASP1

The sample composition for wild type *chicken* Brain abundant signal protein 1 (BASP1, 244 residues; M_w=25438.2; theoretical p_I: 4.66) was 1.4 mM protein, 100 mM sodium phosphate, 150 mM NaCl, pH 6.5, in 90%H_2O and 10%D_2O. The preparation was provided in cooperation by the Klaus Bister group. The primary sequence is given in figure 1.3 (first row).

[1] Dr. Simon Ruedisser, Novartis International, CH-4002 Basel, Switzerland

[2] Prof. Klaus Bister, Institute of Biochemistry, University of Innsbruck, Innrain 52a, A-6020 Innsbruck Austria

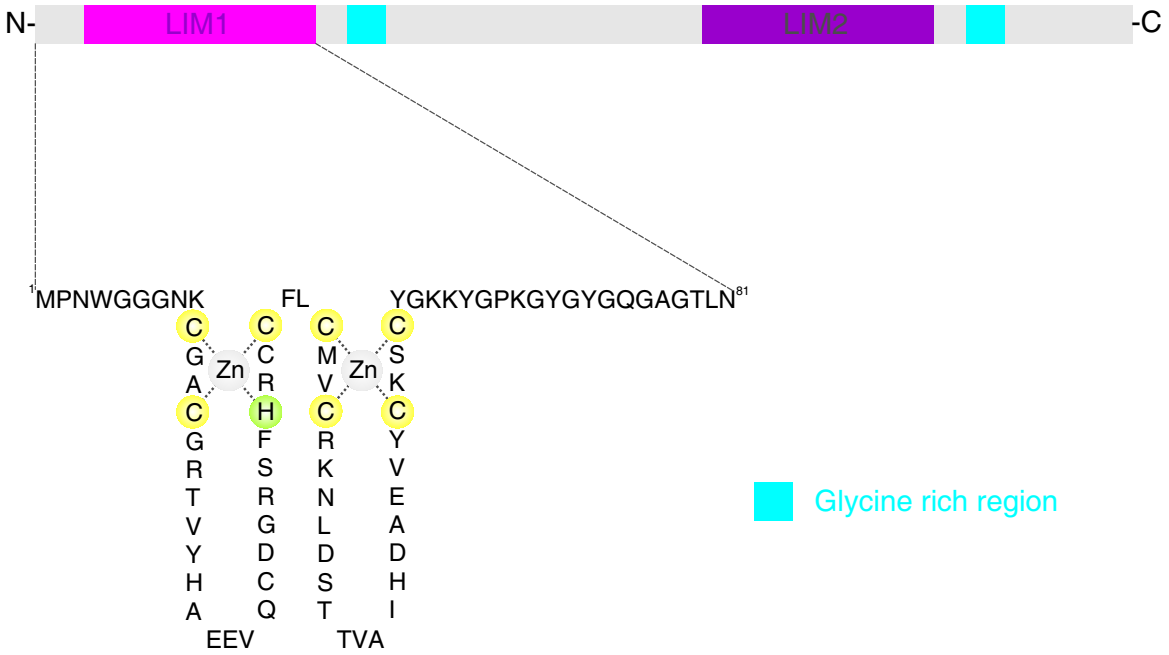


Figure 3.4: Primary sequence of the LIM1 domain containing truncated CRP2 sample used for NMR analysis. The zinc coordinating cysteine and histidine residues are colored in yellow and green, respectively.

3.1.5 Osteopontin

A sample encompassing the C-terminal 219 residues of wild type *quail* osteopontin (OPN, 264 residues; M_w=29335.1; theoretical p_I: 4.54) contained 4.0 mM protein, 100 mM sodium phosphate, 150 mM NaCl, pH 6.5, in 90%H_2O and 10%D_2O. The preparation was provided in cooperation by the Klaus Bister group. The complete wild type primary sequence is given in figure 1.4 (first row).

3.1.6 Materials

Chemical/Material	**Source**
Adenosine-5-monophosphate	Merck, p.a.
L-Argenine.hydrochloride	Sigma, p.a.
Calcium chloride.dihydrate	Fluka p.a.
Centriprep 10 ultrafiltration filters	Amicon Co.
3[(3-Cholamindopropyl)dimethylammonio]-1-propanesulfonate	Sigma, p.a.
Deuterium oxide	Cambridge Isotope Lab.
Dipotasiumhydrogenphosphate.dihydrate	Fluka p.a
D-myo-Inositol 3,4,5-tris-phosphate.tripotassium salt	Alexis & Co

Chemical/Material	Source
DL-myo-Inositol 1,4,5,6-tetrakis-phosphate.tetrapotassium salt	Alexis & Co.
Integrin peptide (M987-WKVGFFKRN-R997)	Fuerst & co-workers
3-(N-morpholino)propanesulfonic acid	Sigma, p.a.
Potasiumchloride	Fluka p.A
Potasiumdihydrogenphosphate	Fluka p.A
Shigemi tubes	Shigemi Co.
Sodiumchloride	Fluka p.A

3.1.7 Experimental NMR parameter

NMR experiments were performed on a Varian Inova 800 MHz (operating at 799.90 MHz for ^{1}H, 201.16 MHz for ^{13}C, and 81.06 MHz for ^{15}N), 600 MHz (599.89 MHz for ^{1}H, 150.88 MHz for ^{13}C, and 60.79 MHz for ^{15}N), and 500 MHz (499.81 MHz for ^{1}H, 125.71 MHz for ^{13}C, and 50.66 MHz for ^{15}N) spectrometer equipped with pulsed field gradient units and triple resonance H/N/C probes. Unless otherwise noted all spectra were acquired at 25°C. Typical carrier positions employed in the double and triple resonance experiments were 119.9 ppm for ^{15}N, 178 ppm for ^{13}C', 58 ppm for $^{13}C^{\alpha}$, 43 ppm for $^{13}C^{\alpha}$/$^{13}C^{\beta}$ (i.e. aliphatic carbons), and 4.68 ppm (H_2O) for ^{1}H. Transmitter offsets and spectral width are chosen to minimize spectral overlap due to aliased signals while maximize resolution. ^{1}H pulse width calibration was done by measurement of a 2π pulse. Indirect pule calibration schemes were applied for the ^{13}C and ^{15}N heteronuclei. Experiments in which protons directly bound to nitrogen were detected during acquisition employed WALTZ ^{15}N decoupling, whereas experiments which detected aliphatic protons during acquisition made use of GARP for ^{13}C decoupling. Where necessary (e.g. HNCACB), carbonyl decoupling was achieved using a SEDUCE-1 ^{13} decoupling sequence centered at 178 ppm. TOCSY spectra (3D H(CCCO)NH and 3D (H)CC(CO)NH) were acquired using the DIPSI-2 sequence [154] for isotropic mixing. All triple resonance experiments performed included a water flip back pulse for minimizing interference arising from radiation damping or solvent. In some of the acquired NOESY experiments a WATERGATE sequence [140] was used for solvent suppression. Quadrature detection in the indirect dimensions was performed using States-TPPI [109] method. However, for ^{15}N evolution the Rance-Kay technique [79] for sign discrimination was employed. ^{1}H chemical shifts were referenced to 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS, δ^{1}H = 0.0 ppm). For heternuclei (^{13}C,^{15}N) indirect calibration

of their chemical shifts was performed using values of 0.25144952 and 0.10132905 for the ratios $\gamma_{^{13}C}/\gamma_H$ and $\gamma_{^{15}N}/\gamma_{^1H}$, respectively. Cross peaks with resonance frequencies outside the spectral width are aliased into the spectrum. Their chemical shifts were corrected according $\delta_{corrected}[ppm] = \delta_{folded}[ppm] \pm n \cdot SW(X)[Hz]/\omega_{lamor}$, were X is either ^{15}N or ^{13}C nuclei. All NMR data were processed using the NMRPipe software package [29], the resulting spectra were assigned and analyzed using the NMRView [71] and SPARKY [48] programs, respectively.

2D ^{1}H,^{15}N HSQC					
Sample	ICln*	Cyc D	LIM1	BASP1	OPN
Spec. width (F_1xF_2)[Hz]	2200x1300			2200x1300	
# of increments F_1	128				
# of scans/increment	8	4	8	4	4
Decoupling scheme	WALTZ-16				
Data size	2x128x832			2x256x512	
Data processing					
Time domain correction	Low frequency filter				
Zero filling	F_1, F_2				
Apodization function	F_1, F_2: squared sine bell, offset 35°				
Baseline correction	F_1: Automated baseline correction				

* sample A to E

2D ^{15}N T_1-experiment					
Sample	ICln (sample B)	ICln (sample E)	LIM1	BASP1	OPN
Spec. width (F_1xF_2)[Hz]	2200x13000			2000x10000	
# of increments F_1 (^{15}N)	128			256	
# of scans/increment	8	4	8	4	4
Decoupling scheme	GARP, WALTZ-16				
Recycle delay [s]	2				
Relaxation delays [ms]	see section 3.3				
Data size	2x128x832			2x256x1024	
Data processing					
Time domain correction	Low frequency filter				
Zero filling	F_1, F_2				
Apodization function	F_1: squared sine bell (first point scaled to 0.5), offset 35° F_2: squared sine bell (first point unscaled), offset 35°				

2D ^{15}N T_2-experiment				
Sample	ICln*	LIM1	BASP1	OPN
Spec. width (F_1xF_2)[Hz]	2200x13000		2000x10000	
# of increments F_1 (^{15}N)	128		256	
# of scans/increment	8			
Decoupling scheme	GARP, WALTZ-16			
Relaxation delays [ms]	see section 3.3			
CPMG delay [ms]	0.5			
γB_1 (^{15}N) [kHz]	3.4			
Data size	2x128x832		2x256x512	
Data processing				
Time domain correction	Low frequency filter			
Zero filling	F_1, F_2			
Apodization function	F_1: squared sine bell (first point scaled to 0.5), offset 35° F_2: squared sine bell (first point unscaled), offset 35°			

* sample B and E

2D ^{1}H,^{15}N NOE-experiment			
Sample	C-term ICln (sample E)	BASP1	OPN
Spec. width (F_1xF_2)[Hz]	3000x13000	2000x10000	
# of increments F_1 (^{15}N)	80		
# of scans/increment	64	32	
Decoupling scheme	GARP, WALTZ-16		
Recycle delay [s] with/without presat.	4/5.5		
Presaturation delay [s]	1.5		
Data size	2x80x408	2x80x512	
Data processing			
Zero filling	F_1, F_2		
Apodization function	F_1: squared sine bell (first point scaled to 0.5), offset 35° F_2: squared sine bell (first point unscaled), offset 35°		
Baseline correction	F_1: Automated baseline correction		

2D ^{15}N spin echo difference CT HSQC		
Sample	ICln (sample B)	LIM1
Spec. width (F_1xF_2)[Hz]	6000x9000	
# of increments F_1 (^{15}N)	220	
# of scans/increment	128	
Decoupling scheme	GARP, WALTZ-16	
Constant time period T [ms]	29.8	
Data size	2x2x220x288	
Data processing		
Zero filling	F_1, F_2	
Apodization function	F_1, F_2: squared sine bell, offset 35°	
Baseline correction	F_1: Automated baseline correction	

2D ^{13}C spin echo difference CT HSQC		
Sample	ICln (sample B)	LIM1
Spec. width (F_1xF_2)[Hz]	6000x9000	
# of increments F_1 (^{15}N)	220	
# of scans/increment	128	
Decoupling scheme	GARP, WALTZ-16	
Constant time period T [ms]	29.8	
Data size	2x2x220x288	
Data processing		
Zero filling	F_1, F_2	
Apodization function	F_1, F_2: squared sine bell, offset 35°	
Baseline correction	F_1: Automated baseline correction	

2D NOESY	
Sample	ICln (sample B)
Spec. width (F_1xF_2)[Hz]	13000x13000
# of increments F_1 (^{15}N)	256
# of scans/increment	128
Decoupling scheme	WURST-140
Mixing time τ_{mix} [ms]	150
Data size	2x256x833
Data processing	
Low frequency substraction	Polynom 6^{th} order
Zero filling	F_1, F_2
Apodization function	F_1, F_2: squared sine bell, offset 35°
Baseline correction	F_1, F_2: Automated baseline correction

3D HNCO					
Sample	ICln	Cyc D	LIM1	BASP1	OPN
Spec. width (F_1xF_2xF_3)[Hz]	3020x8000x2130			2000x10000x2000	
# of increments F_1 (^{13}C)	32			60	
# of increments F_3 (^{15}N)	32			58	
# of scans/increment	16			2	
Decoupling scheme	DIPSI-2(^{1}H), WALTZ-16(^{15}N)				
Data size	2x32x32x512			2x58x60x512	
Data processing					
Time domain correction	Low frequency filter				
Low frequency subtraction	Polynom 6th order				
Zero filling	F_1, F_2, F_3				
Apodization function	F_1,F_3: sine bell (first point scaled to 0.5), offset 35° F_2: squared sine bell (first point unscaled), offset 35°				

3D HN(CA)CO			
Sample	C-term ICln (sample E)	BASP1	OPN
Spectral width (F_1xF_2xF_3)[Hz]	3020x8000x2130		
Number of increments F_1 (^{13}C)	32	64	
Number of increments F_3 (^{15}N)	32	45	
Number of scans/increment	16	16	16
Decoupling scheme	DIPSI-2(^{1}H), WALTZ-16(^{15}N)		
Data size	2x32x32x512	2x64x45x512	
Data processing			
Time domain correction	Low frequency filter		
Zero filling	F_1, F_2, F_3		
Apodization function	F_1,F_3: sine bell (first point scaled to 0.5), offset 35° F_2: squared sine bell (first point unscaled), offset 35°		

3D HNCA					
Sample	ICln	Cyc D	LIM1	BASP1	OPN
Spectral width (F_1xF_2xF_3)[Hz]	3020x8000x2130			3020x8000x2130 ??	
Number of increments F_1 (^{13}C)	32			64	
Number of increments F_3 (^{15}N)	32			45	
Number of scans/increment	16			4	
Decoupling scheme	DIPSI-2(^{1}H), WALTZ-16(^{15}N)				
Data size	2x32x32x512				
Data processing					
Time domain correction	Low frequency filter				
Zero filling	F_1, F_2, F_3				
Apodization function	F_1,F_3: sine bell (first point scaled to 0.5), offset 35° F_2: squared sine bell (first point unscaled), offset 35°				

3D HN(CO)CA			
Sample	C-term ICln (sample E)	BASP1	OPN
Spectral width (F_1xF_2xF_3)[Hz]	3020x8000x2130	2x?x?x?	
Number of increments F_1 (^{13}C)	32	60	
Number of increments F_3 (^{15}N)	32	45	
Number of scans/increment	16	8	
Decoupling scheme	DIPSI-2(^{1}H), WALTZ-16(^{15}N)		
Data size	2x32x32x512	2xxx?	
Data processing			
Time domain correction	Low frequency filter		
Zero filling	F_1, F_2, F_3		
Apodization function	F_1,F_3: sine bell (first point scaled to 0.5), offset 35° F_2: squared sine bell (first point unscaled), offset 35°		

3D HNCACB					
Sample	ICln	Cyc D	LIM1	BASP1	OPN
Spectral width (F_1xF_2xF_3)[Hz]	8000x8000x1750			10000x10000x2000 ?	
Number of increments F_1 (^{13}C)	32			64	
Number of increments F_3 (^{15}N)	45			45	
Number of scans/increment	16			8	
Decoupling scheme	DIPSI-2(^{1}H), WALTZ-16(^{15}N)				
Data size	2x45x32x512				
Data processing					
Time domain correction	Low frequency filter				
Zero filling	F_1, F_2, F_3				
Apodization function	F_3: sine bell (1st point scaled to 0.5), offset 35° F_1, F_2: sq. sine bell (1st point unscaled), offset 35°				

3D CBCA(CO)NH					
Sample	ICln	Cyc D	LIM1	BASP1	OPN
Spectral width (F_1xF_2xF_3)[Hz]	7500x8000x1500			10000x12000x2000 ?	
Number of increments F_1 (^{13}C)	40			64	
Number of increments F_3 (^{15}N)	40			55	
Number of scans/increment	16			8	
Decoupling scheme	DIPSI-2(^{1}H), WALTZ-16(^{15}N)				
Data size	2x40x40x512				
Data processing					
Time domain correction	Low frequency filter				
Zero filling	F_1, F_2, F_3				
Apodization function	F_3: sine bell (1st point scaled to 0.5), offset 35° F_1, F_2: sq. sine bell (1st point unscaled), offset 35°				

3D TOCSY-HSQC	
Sample	ICln (sample B)
Spectral width (F_1xF_2xF_3)[Hz]	8000x10000x2000
Number of increments F_1 (^{1}H)	80
Number of increments F_3 (^{15}N)	40
Number of scans/increment	16
Decoupling scheme	DIPSI-2(^{1}H), WALTZ-16(^{15}N)
Data size	2x80x40x512
Data processing	
Time domain correction	Low frequency filter
Zero filling	F_1, F_2, F_3
Apodization function	F_1, F_3: squared sine bell, offset 35°
	F_2: sine bell, offset 50°

3D ^{13}C,^{15}N NOESY-HSQC			
Sample	ICln	Cyc D	LIM1
Spectral width (F_1xF_2xF_3)[Hz]	8100x12000x2110		
Number of increments F_1 (^{1}H)	55		
Number of increments F_3 (^{15}N)	50		
Number of scans/increment	8		
Mixing time τ_{mix} [ms]	150		
Decoupling scheme	WALTZ-16(^{15}N)		
Data size	2x55x50x512		
Data processing			
Time domain correction	Low frequency filter		
Linear prediction	F_1, F_3		
Zero filling	F_1, F_2, F_3		
Apodization function	F_1, F_3: sine bell (1st point scaled to 0.5), offset 35°		
	F_2: sq. sine bell (1st point unscaled), offset 35°		

3D H(CCCO)NH, 3D (H)CC(CO)NH			
Sample	ICln	Cyc D	LIM1
Spectral width (F_1xF_2xF_3)[Hz]	8000x10000x1750		
Number of increments F_1 (^{1}H)	45		
Number of increments F_3 (^{15}N)	32		
Number of scans/increment	32		
Decoupling scheme	DIPSI-2(^{1}H), WALTZ-16(^{15}N)		
Mixing time τ_{mix} [ms]	75		
Data size	2x55x50x512		
Data processing			
Time domain correction	Low frequency filter		
Linear prediction	F_1, F_3		
Zero filling	F_1, F_2, F_3		
Apodization function	F_2, F_3: sine bell (1st point scaled to 0.5), offset 35° F_1: squared sine bell (1st point unscaled), offset 35°		

3D HNN, 3D HN(C)N		
Sample	BASP1	OPN
Spectral width (F_1xF_2xF_3)[Hz]	2000x10000x2000	
Number of increments F_1 (^{1}H)	45	
Number of increments F_3 (^{15}N)	45	
Number of scans/increment	8	
Decoupling scheme	GARP(^{1}H), WALTZ-16(^{15}N)	
Data size	2x45x45x512	
Data processing		
Time domain correction	Low frequency filter	
Linear prediction	F_1, F_3	
Zero filling	F_1, F_2, F_3	
Apodization function	F_2, F_3: sine bell (1st point scaled to 0.5), offset 35° F_1: squared sine bell (1st point unscaled), offset 35°	

3.2 NMR derived structural restraints

3.2.1 Distance restraints

Interproton distances were inferred from NOE cross peaks in the 3D ^{15}N- and ^{13}C edited NOESY-HSQC experiments, respectively. Cross peaks volumes were calculated using NMRView software [71] and calibrated to distances restraints on the basis of characteristic short interproton distances [126] found in elements of regular secondary structure (e.g. $d_{H_i^N-H_i^N}$, $d_{H_i^N-H_{i-3}^\alpha}$ in α-helices or $d_{H_i^N-H_i^\alpha}$, $d_{H_i^N-H_{i-1}^\alpha}$ in β-strands) or classified as strong (1.8-2.2Å), medium (1.8-3.4Å), and weak (1.8-6Å). For nonstereospecifically assigned or equivalent protons (methylene, methyl and aromatic hydrogens), NOE distances were implemented as a $\sum r^6$ sum [41] into structure calculation.

3.2.2 Hydrogen bonds

Hydrogen bonds were derived from NOE pattern analysis and backbone secondary chemical shifts. In parallel and anti-parallel β-sheets a backbone amide hydrogen bond was defined whenever inter-β-strand ^{1}H$_i^N$-^{1}H$_j^N$ NOE contacts to a residue j bearing the acceptor carbonyl C' and to the residue $i+2$ (in anti-parallel β-sheets) or $i-2$ (in parallel β-sheets) were observed. Additionally, long-range ^{1}H$_i^N$-^{1}H$_j^\alpha$ (in parallel and anti-parallel β-sheets) and the strong ^{1}H$_i^\alpha$-^{1}H$_j^\alpha$ NOE correlation found in anti-parallel β-sheets were utilized for hydrogen bond identification (see figure 3.5.A and 3.5.B). In regions of regular α-helical secondary structure, a hydrogen bond was defined between the carbonyl function C' of residue i and the backbone amide proton H^N of residue $i+4$ (figure 3.5.C).

Each hydrogen bond was explicitly represented by two distance restraints $d_{N_i-H_i}$ and $d_{N_i-O_j}$ in order to favor a roughly linear arrangement of the atoms $N_i - H_i \cdots O_j$ (where i and j denotes the sequential positions of the atoms). These restraints were set to 1.8Å $< d_{N_i-H_i} <$ 2.1Å and 2.5Å $< d_{N_i-O_j} <$ 3.1Å, respectively.

3.2.3 Dihedral angle restraints

3.2.3.1 Backbone φ,ψ predictions from the TALOS program

Using the assigned ^{15}N, ^{13}C$^\alpha$, ^{1}H$^\alpha$, ^{13}C$^\beta$ and ^{13}C' chemical shifts, predictions for the backbone ϕ,ψ torsion angles were obtained by using the TALOS program [25].

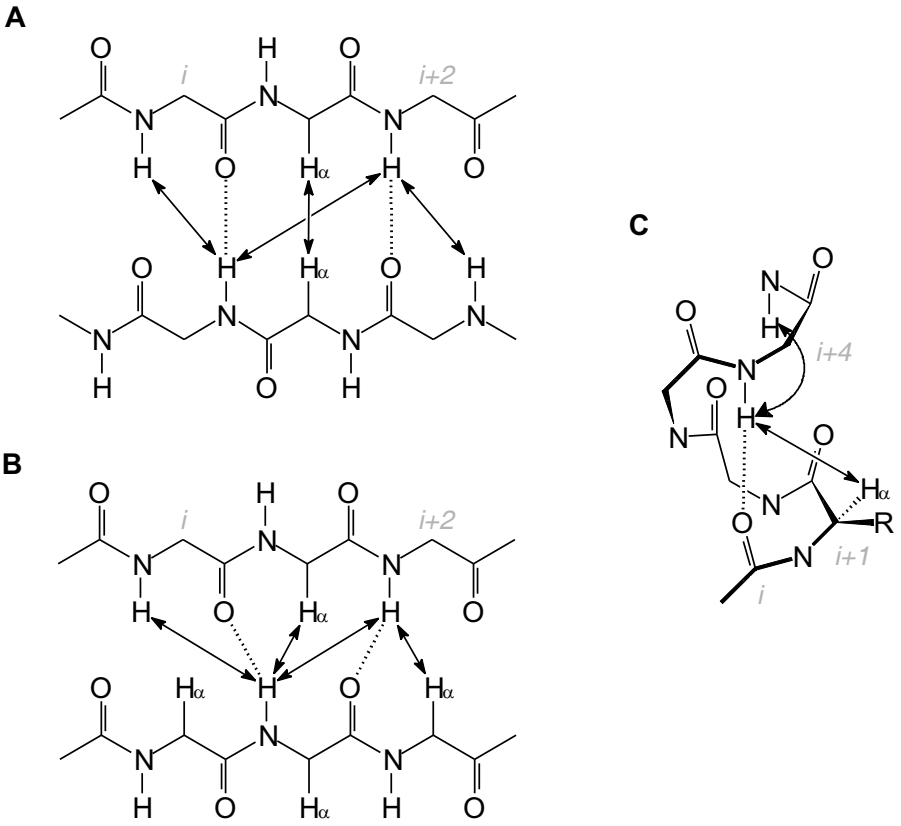


Figure 3.5: Characteristic NOE contacts found in protein secondary structure elements.

At the initial stage of structure calculation, only predictions were accepted and classified as *good*, for which at least 8 out of 10 of the best database matches fall in or near the same (empirical well-populated) region of the Ramachandran map. The average values Θ_{avg} (with Θ_{avg} = ϕ_{avg} or ψ_{avg}) of these predictions associated with a standard deviation σ_Θ were consequently converted to dihedral angel restraints and implemented in subsequent structure calculations using $\Theta_{avg} + \sigma_\Theta$ and $\Theta_{avg} - \sigma_\Theta$ as upper and lower bounds, respectively.

At later stages of structure determination backbone ϕ,ψ torsion angles predictions, which were initially classified as ambiguous (but not necessarily wrong), were successive added to the list of torsion restraints.

3.2.3.2 Sidechain χ_1 restraints

From ^{13}CO- and ^{15}N spin echo constant time HSQC experiments $J_{C^\gamma C'}$ and $^3J_{C^\gamma N}$ couplings between the backbone carbonyl ^{13}C'and amide nitrogen ^{15}N, and the methyl carbons -$^{13}CH_3$ of valine, isoleucine ($C^{\gamma 2}$), and threonines, respectively, were measured. From this data and

equation 2.7, the rotameric states (g^+, g^-, or t) of the methyl carbons C$^\gamma$ to the intra-residue backbone carbonyl for residues Val35, Thr69, Ile70, Val89, Val103, Ile116, and Val122 were inferred. Subsequent, χ_1 torsion angle restraints were defined using upper and lower bounds of +60° and -60° relative to the target rotameric state, respectively, and implemented for structure calculation in an analog manner as given by equation 2.52.

3.3 Protein dynamic NMR measurements

Dynamic information was obtained by measurement of backbone ^{15}N^N T_1 and T_2 relaxation times as well as heteronuclear steady state NOE values, as described in [39]. Relevant acquisition parameters are given in section 3.1.7.

3.3.1 Backbone ^{15}N relaxation of ICln

For 'truncated' wild type ICln (sample B), relaxation parameters were measured at 25°C and a field strength of 499.86 MHz and 799,89 MHz, respectively. The ^{15}N transverse relaxations experiments for evaluating T_2 were performed with Carr-Purcell-Meiboom-Gill (CPMG) delays of 10, 30, 50, 70, 90, and 110 ms for sample B, and 70, 130, 190, 250, 310, and 370 ms for C-terminal ICln (sample E), respectively, employing a CPMG duty cycle delay of 0.5 ms. ^{15}N longitudinal relaxation T_1 parameter measurements were obtained using delays of 0, 200, 400, 600, 800, and 1000 ms for sample B, and 200, 400, 600, 800, 1000, and 1200 ms for sample E, respectively. The sequence of cross-peaks intensities in that spectra series was fit to a three parameter decaying exponential of the form $I(t) = A\exp[-(t/T)]+C$, where $I(t)$ is peak height and t the delay parameter, using the rate analysis module of the nmrView software package [71] or an extension script of SPARKY [48], respectively. Random errors for T_1 and T_2 values were estimated by calculating a best fit for a set of perturbed heights 5000 times. The heights were perturbed by computer-simulated noise with a Gaussian distribution of zero mean and with variance equal to the root-mean-square deviation of the original heights from the original best fit. The spread in the best fit values was used as a measure of the experimental random error.

Heteronuclear steady state NOE ^{15}N$\{^1$H$^N\}$ attenuation factors for C-terminal ICln (sample E) were derived from the ratio I_{NOE}/I_{nonNOE}, where I_{NOE} and I_{nonNOE} denotes the peak intensities in the experiments with and without proton saturation, respectively [171]. In the case of the spectra with non saturation, a net relaxations delay of 5 s was employed, while a relaxation

delay of 2 s prior to a 3 s proton presaturation period was applied for the NOE spectra. The standard deviation of evaluated ^{15}N$\{^{1}$H$\}$ attenuation values was determined as described in [39].

3.3.2 Backbone ^{15}N relaxation measurements and spectral density analysis for BASP1 and Osteopontin

Backbone ^{15}N^{N} T_1 and T_2 relaxation values were measured according [39] at 25°C and a field strength of 600.13 MHz.

Carr-Purcell-Meiboom-Gill (CPMG) delays in the transverse T_2 relaxations experiments for BASP1 and Osteopontin were set to 0, 16, 32, 64, 128, 192, and 256 ms, respectively, using a delay of 0.5 ms between successive application of ^{15}N 180° and ^{1}H 180° pulses applied every 4 ms in the CPMG pulse train. Longitudinal ^{15}N T_1 relaxation parameter were measured using delays of 0, 20.8, 41.6, 83.2, 208, 416, 832, 1248, and 1664 ms for both proteins. Relaxation data analysis was performed as described above (section ?) using an extension script of SPARKY [48]. Random errors for T_1 and T_2 values were estimated by calculating a best fit for a set of perturbed heights 5000 times. The heights were perturbed by computer-simulated noise with a Gaussian distribution of zero mean and with variance equal to the root-mean-square deviation of the original heights from the original best fit. The spread in the best fit values was used as a measure of the experimental random error.

Heteronuclear steady state NOE ^{15}N$\{^{1}$H$^{N}\}$ attenuation factors and their related standard deviation values were determined as given in section 3.3.1.

3.4 NMR structure determination

Three dimensional structures were generated starting from a covalent template with randomized dihedral angles by a molecular dynamic simulated annealing (md-SA) protocol using the XPLOR-NIH software package [164] including beside energy terms for maintaining the covalent geometry, a simple repulsive quartic van der Waal term, energy terms accounting for the NMR observables, a knowledge based potential of mean force [91] to bias ϕ, ψ and χ_1-torsions to regions that are preferentially populated, and a potential term for the radius of gyration [92] improving the overall packing.

3.4.1 The forcefi eld

In all structure calculations performed the CHARMM forcfield [16]was applied. The data structure in CHARMM consists of a protein structure file (.psf) a protein topology parameter file (.pro), and an atom coordinate file (.pdb). The .psf file was derived from the topology file and the primary sequence of the protein and contained atom information (atom names, types, charges, masses), lists of bonds, dihedral, and improper dihedrals for the energy terms, hydrogen donor and acceptor lists, and non-bounded exclusion list for atom pairs connected by close interactions (1-2, and 1-3; for excluding van der Waal term). The file pair '$parallhdg.pro$' (containing the force constant parameter values) and '$topallhdg.pro$' (including the protein residue topology as well as charge and mass values) was used as protein topology parameter information throughout this work. Where necessary, covalent modifications of the structure template (for disulfide bonds, zinc fingers, or polypeptide chain termination) were introduced using the PATCH command of the CHARMM command parser. As no evidence for cis prolines was found, all peptide bounds of the proteins under investigation were restrained to *trans*.

3.4.2 Generating the structure templates

For the md-SA protocols of ICln and Cyclophilin D a starting (elongated) covalent protein structure templates offering a good local geometry free of bad non-bonded contacts were generated by alternating steps of Powell minimization and 0.05 ps dynamics at 300 K, respectively, until the CHARMM forcefield total energy obtained was below 1000 kcal/mol.

In the case of LIM1, the zinc fingers motifs were initially omitted in order to prevent entanglement of the starting structure. Instead, a pre-minimized elongated LIM1 template structure was generated as described above but now employing the whole set of initial unambiguously defined experimental distance restraints in context with synthetic distance restraints to arrange a roughly tetragonal geometry of the zinc finger atoms. The synthetic data contained artificial distance restraints between all Zn^{2+} coordinating atoms (Zn^{2+} finger I (CCHC, N-terminal domain): Cys10(S^{γ}), Cys13(S^{γ}), His31($N^{\delta 1}$), and Cys34(S^{γ}); Zn^{2+} finger II(CCCC, C-terminal domain) Cys37(S^{γ}), Cys40(S^{γ}), Cys58(S^{γ})), and Cys61(S^{γ})) and between those atoms to the central Zn^{2+} atoms, whereas the target distance values d_{S-S}, d_{S-N},$d_{Zn^{2+}-S}$ and $d_{Zn^{2+}-N}$ were set to 3,4$\pm$0.2 Å, 3.2$\pm$0.2 Å, 2.4$\pm$0.2 Å, and 2.2$\pm$0.2 Å, respectively.

Then, the final template for subsequent structure refinement was obtained by covalent linking of the Zn^{2+} atoms to its coordinating cysteine and histidine residues.

3.4.3 Molecular dynamic simulating annealing protocol

Except for the cooling phase, which was chosen shorter for the structure calculation of LIM1, identical setup values were employed for all proteins.

Initial atom velocities were allocated according a Maxwell distribution at 2000 K. All atomic masses were set to uniform values (100 a.u.). The first phase of the protocol includes 10 ps (10000 steps) dynamical simulated annealing using a time step of 1 fs at 2000 K. The force constant values $k_{torsion}$ and k_{NOE} of the torsion- and NOE energy terms were chosen to 2 kcal/mol·Å^2, whereas the force constants of the potential terms specifying the covalent geometry and atomic repulsion were adjusted to significant low values (0.004 kcal/mol·Å^4 for k_{vdW}, 0.4 kcal/mol·Å^2 for the k_{angles}, and 0.1 kcal/mol·Å^2 for $k_{improper}$). Values of 5 kcal/mol·Å^2 and 0.002 kcal were set for the weighing factors of the radius of gyration pseudopotential $k_{gyration}$ and the empirical database potential k_{rama}, respectively. Controlling of the temperature was achieved by the Berendsen method [110].

At the second stage of the protocol comprising 40 cycles of 2.25 ps dynamics (30000 steps, time step 3 fs) in the setup of ICln and cyclophilin D, and 40 cycles of 0.75 ps dynamics (10000 steps, time step 3 fs) in the case of LIM1, respectively, all force constants values were successively increased by multiplication with factors according equation 3.1 up to maximum values of 30 kcal/mol·Å^2 for k_{NOE}, 200 kcal/mol·Å^2 for $k_{torsion}$, 1.2 kcal for k_{rama}, 20 kcal/mol·Å^2 for $k_{gyration}$, 4 kcal/mol·Å^4 for k_{vdW}, and 1 kcal/mol·Å^2 for all geometry related force constants.

$$(v_{final}/v_{intitial})^{1/n_{cycles}} \tag{3.1}$$

where $v_{intitial}$ and v_{finial} are the starting and final force constant values, respectively, at which the number of cycles, n_{cycles}, is given by $(T_{final} - T_{intitial})/\Delta t$, whereas $T_{intitial}$ and T_{final} denotes the initial and final temperature values, respectively, and Δt the choosen time step.

The temperature was decreased down to 10 K. A graphical summary of the force constant scaling is given in figure 3.6.

Finally, a restrained Powell minimization [18] was applied whose maximal number of minimization cycles was set to 500. A gradient tolerance (used for identification of local minima or saddle points) of 0.0001 was used. In course of progressive steps of structure calculation and analysis, further restraints were classified to be unambiguous and introduced into the database of restraints.

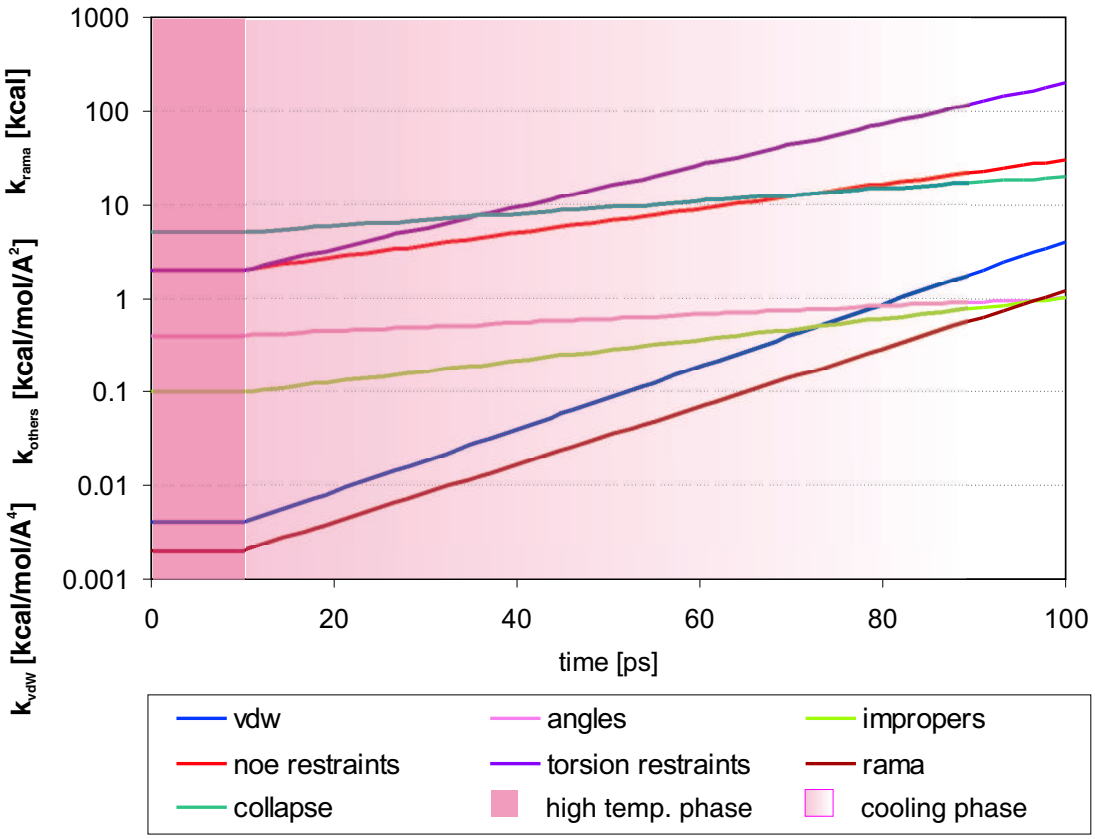


Figure 3.6: Scaling of the force constants k_{NOE} [kcal/mol·Å^2], $k_{torsion}$ [kcal/mol·Å^2], k_{angles} [kcal/mol·Å^2], $k_{improper}$ [kcal/mol·Å^2], k_{rama} [kcal], $k_{gyration}$ ($k_{collapse}$) [no dimension], and k_{vdW} [kcal/mol·Å^4] related to distance restraints, torsions, angles, improper dihedrals, empirical database potential for $\phi/\psi/\chi_1$- torsions, and to the radius of gyration, respectively, in the applied molecular annealing protocol.

3.4.4 Validation of the obtained structure models

At the end of each structure determination, a final bundle of 15 protein models with a minimum number of restraints violations and minimal total energy were selected for structural validation. The extent of experimental distance and dihedral angle restraint violations were computed by using the AQUA program [96]. Assessment of the backbone and sidechain torsion angles and visualization of the validation results produced by AQUA were performed using the PROCHECK-NMR software package [96]. The qualities of the backbone ϕ,ψ- and sidechain χ_1,χ_2 torsion angle distributions were verified as described by [107]. Correctness of bond lengths, improper dihedrals for maintaining covalent chirality and planarity, peptide bond ω torsion angles and the packing environment was validated using the protein check subroutine of the WHAT IF web interface (http://swift.cmbi.kun.nl/WIWWWI/).

To further characterize the obtained structure model qualities of the proteins studied against a empirical database potential (translated from a SCOP release from September 2002) the program PROSA2003 was used.

Unless otherwise noted the energy graphs for structural diagnostics produced in this work arise from a combined C^{α},C^{β} atom pair $\Delta P(S,C)$ and surface $\Delta S(S,C)$ energy. The weighting factors for both energy terms ω_p and ω_s (see equation 2.65) were set to 1 and according the ratio $\sigma_{pair}/\sigma_{surface}$ (default values), respectively, whereas σ_{pair} and $\sigma_{surface}$ denotes the standard deviation of the atom pair and surface energy distribution, respectively.

The distance cutoff for calculation of the atom pair energies $\Delta P(S,C)$ was set to 4-15Å (outside that range the energies are zero), thus potential energy distribution arising from close contacts or other violations of basic steric principles are here not taken into account (as they still analyzed using the WHATIF web routine). Virtual glycine C^{beta} atoms, as optionally provided by PROSA2003, are omitted for all energy calculations. Unless otherwise noted, a gliding average window of 10 residues was applied for the representation of the residual energy profile, considering the 5 adjacent residues at the N-terminal and C-terminal extremes. For calculation of the z-score values the polyprotein $pII3.0.short.ply$ was used as conformation template.

The precision of the final structure bundles in terms of RMSD and circular variance values was calculated using the programs MOLMOL [85] and PROCHECK [96], respectively.

3.5 Towards automated protein fold generation

3.5.1 Semi-automated protein structure calculation with a priori sequential backbone resonance assignment

To illustrate the benefit of selective protonation techniques in context with automated methyl chemical shift assignment to protein fold determination, structure calculations for the proteins *dog* ICln (168 residues), *human* Cyclophilin D (165 residues) and *quail* lipocalin Q83 (157 residues) by using either pure experimental or synthetic NOE data were performed. The structures of ICln and Cyclophilin D are described in section 4.1.1 and 4.1.2. The lipocalin Q83 protein consists of a eight stranded up and down β-barrel flanked at the C-terminal side by an 13 residue α-helix. Starting from a linear (pre-minimized) template the employed calculation protocol (including empirical database potential for ϕ, ψ and χ_1 torsions and a pseudo-potential for the radius of gyration) used the same simulated annealing parameter as described in sec-

tion 3.4.3. To visualize the influence of the methyl group related NOEs in respect to attainable fold quality, the obtained structures were compared to those derived using backbone H^N-H^N NOEs as the sole type of distance restraints.
The input data contained:

- NOE distances In case of ICln and Cyclophilin D exclusive experimentally observed NOE distances were used for structure calculation including backbone H^N-H^N, backbone to sidechain H^N-CH_3^{ILV} and sidechain CH_3^{ILV}-CH_3^{ILV} NOE contacts, whereas CH_3^{ILV} denotes a methyl group of either leucine ($H^{\delta 1}$,$H^{\delta 2}$), valine ($H^{\gamma 1}$,$H^{\gamma 2}$), or isoleucine ($H^{\delta 1}$ only) residue type. Backbone amide NOE distances were roughly classified according their NOE cross-peak volumes (weak, medium and strong; see section 3.2.1), the upper bound of other distance restraints were uniformly set to 6Å.

  For lipocalin Q83, synthetic backbone H^N-H^N, backbone to sidechain H^N-CH_3^{ILV} and sidechain CH_3^{ILV}-CH_3^{ILV} NOE distances were derived from the structure coordinates deposited at the protein data bank (PDB entry code: 1JZU) using a spatial distance cutoff values c_{off} of 6Å (a NOE between two protons was assumed, whenever their spatial distance was shorter then c_{off}). The extent of the NOE databases for the proteins are given below in table 3.19.

- Hydrogen bonds As readily obtainable from backbone NOE pattern analysis, experimentally inferred hydrogen bonds for ICln and Cyclophilin D were used for the structure calculations. For lipocalin Q83 the hydrogen bonds were calculated from the deposited structure (1JZU) using the program MOLMOL (version 2K.2 [86]). A hydrogen bond between a backbone amide hydrogen H^N and a carbonyl oxygen O acceptor was defined whenever their distance was shorter then 2.4Å and the included angle between the vectors $\overrightarrow{H^NO}$ and $\overrightarrow{NO}$ was smaller then 50°. However, increased upper bounds b_{upper} of 3.3Å and 2.5Å for the d_{H^NO} and d_{NO} hydrogen bond related distance restraints were employed for the structure calculation of Q83 to account for their synthetic origin. For the experimentally derived proteins ICln and Cyclophilin D the original b_{upper} parameter values were used.

- Backbone ϕ,ψ torsion restraints from TALOS predictions Backbone shifts of lipocalin Q83 protein were taken from the BMRB (BioMagResBank, www.bmrb.wisc.edu), accession number 4664. For the other proteins their experimentally derived backbone shifts were used for TALOS ϕ,ψ torsion angle prediction.

The total number of restraints employed for structure calculation are shown in table 3.19.

Table 3.19: Extent of applied NOE distance (backbone H^N-H^N and sidechain H^N-CH_3(I,L,V)/CH_3(I,L,V)-CH_3(I,L,V)) and torsion restraints applied for minimalistic structure calculation protocol. Number of backbone H^N-H^N contacts are given in brackets.

Protein	**ICln**	**Cyclophilin D**	**lipocalin Q83**
intra-residue	12	50	80
sequential $\|i-j\|$ = 1 (bb)	107 (26)	170 (46)	183 (40)
medium-range 4 $\geq \|i-j\| \geq$ 2 (bb)	37 (19)	63 (21)	85 (13)
long-range $\|i-j\| >$ 4 (bb)	124 (81)	150 (81)	207 (107)
Total NOE distances	270 (138)	434 (198)	315 (240)
Hydrogen bonds N-H$\cdots$O-C	48	58	?
bb ϕ,ψ torsions (TALOS)	98	80	114

Because the orientation of the α-helix in the lipocalin Q83 protein (residue Ala124 to residue Arg136) is partly determined by a disulfide bridge between Cys59 and Cys152, two sets of structure ensemble with and without that covalent linkage was calculated for comparison purposes.

3.5.2 Rapid protein fold validation using a direct method in context with residue type specifi c isotope labeling

The direct method procedure described in chapter 2 was tested for the structure generation of three experimentally derived proteins, Cyclophilin D (165 residues) [158], ICln (168 residues) [159], and LIM1(CRP2)(86 residues)(data not published, see section 4.1.3), which were originally manually obtained at this work in the course of conventional (indirect) structure determination analysis. The derived NMR distance restraints used here were exclusively generated from unambiguously assigned NOESY cross peak lists available from those conventional structure determinations.

The validate the performance of the method for other different protein fold topologies idealized NOE data sets encompassing distances comprising H^N and CH_3(ILE,LEU,VAL) atoms only were derived from the PDB coordinates of four other proteins displaying diverse secondary and tertiary structures by using an interatomic distance cutoff of 6Å, within NOE cross peaks were assumed to be observed. The folds of the experimentally derived Cyclophilin D (165 residues), ICln (168 residues), and LIM1 domain of the *quail* cysteine rich protein CRP2 (86 residues) are described in sections 4.1.1 to 4.1.3, respectively. Calmodulin (148 residues) from *Drosophila melanogaster* consists of a seven-turn central α-helix connecting two calcium-

binding α-domains. *Human* Carbonic Anhydrase I (260 residues) contains a saddle-shaped ten-stranded antiparallel β-sheet covered at the top by a pair of α-helices, two 3/10-helices, and one anti-parallel β-sheet, whereas to bottom side is faced by a small three-stranded parallel/ antiparallel-β-sheet, one α-helix and two additional $3/10$-helices. Plastocyanin (99 residues) from *French bean* adopts a beta-sandwich structure composed of two four-stranded parallel/antiparallel-β-sheets. Human Thioredoxin (105 residues) consists of a central five-stranded antiparallel-β-sheet surrounded by four α-helices.

3.5.2.1 Structure Calculations: Simulated Annealing I

An ensemble of spatial atomic distributions was calculated via MD-Simulated Annealing starting from 300 independent templates lacking any covalent linkage (round 0), each initially randomized within a cube of side length l=50 Å except for the larger-sized *human* carbonic anhydrase I, were l=100Å was necessary for an optimal minimization performance. Only an attractive soft square-well potential U_{NOE} based on NOE distances and a repulsive Van der Waal term U_{VdW} were active in the applied energy forcefield. The high temperature phase included 10 ps molecular dynamic at 4000 K using a time step of 1fs. The initial value for the force constant f_{NOE} and f_{VdW} were 2 kcal/mol·Å^2 and 4 cal/mol·Å^4, respectively, and the ratio $f_{NOEPseudo}/f_{NOE}$ was set to 1:100. During cooling stage, which comprised 80 cycles of 3.75 ps dynamics for *human* carbonic anhydrase I and 80 cycles of 1.125 ps dynamics for all other proteins (timestep 3fs), the temperature was lowered to 10 K and all forceconstants were changed to their final values (f_{NOE} = 300 kcal/mol·Å^2, $f_{NOEPseudo}$ = 3 kcal/mol·Å^2, and f_{VdW} = 4 kcal/mol·Å^4). Finally 1000 steps of Powell energy minimization were applied.. All resulting structures displayed not any input NOE distance restraint violations.

3.5.2.2 Statistical Analysis of Conserved Distances

Analysis or the resulting atomic density clouds according section 2.7.4 on page 107 was performed using home written perl and .tcl scripts. Structural superposition and visualization was done using MOLMOL version 2k.2 [85] or vmd-xplor version 1.4 [164].

3.5.2.3 Simulated Annealing II and Statistical Analysis II

The artificial distance restraints A^1_{ij} derived from the initial atomic distribution were merged with the experimental NOE distances and used for a subsequent calculation (iteration round 1)

performed in an analogical manner as described above. However, the forceconstant $f_{artificial}$ of A_{ij} were set to a fifth part relative to f_{NOE} to ensure that all original NOE distances remain fulfilled. A set of 300 atomic distributions was calculated, from which the best 50 structures (in respect to lowest U_{NOE} energy values) were taken for subsequent determination of statistical conserved distances. The presence of artificial distance restraints A_{ij} in the structure calculation giving raise to small σ_{ij} values of distances r_{ij} which are not necessarily adequate defined by experimentally derived NOEs, thus potentially not compatible with the correct structure. Because of this the more stringent criteria listed on page ? are used for conversion of statistical conserved distances to new artificial distance restraints A^2_{ij}. The upper and lower bounds of defined artificial distance restraints defined in previous rounds were let unchanged, since (despite more strict selection) there was empirically a remaining small bias to larger distances observed, which increases slowly in subsequent rounds of structure calculations.

3.5.2.4 Model Driven Sequential Chemical Shift Assignment

The benefit of the obtained non-covalent structures in context with automated backbone chemical shift assignment were examined by using the simulated annealing Monte Carlo (SA-MC) program MONTE [63], although other MC programs should comply the purposes as well (e.g. MARS [72], or [106]).

By default, MONTE dry to recalculate the backbone amide H_N proton positions of the input structure used for NOE pattern matching, so a semicovlent template providing the vectors $\overrightarrow{C'_i N_i}$ and $\overrightarrow{C_{\alpha,i} N_i}$ were needed. For that purpose the last calculation step (iteration round2) was repeated to obtain a .pdb file with covalently bonded backbone C'_{i-1}, N_i, H^N_i, and $C_{\alpha,i}$ atoms, but lacking the intraresidue $C_{\alpha,i}$-C'_i bond to maintain the covalent in-dependency of these C'_{i-1}-N_i(H^N_i)-$C_{\alpha,i}$ fragments to each other.

The provided input information for automated backbone assignment included the primary sequence of the proteins, intra- and inter-residue carbon chemical shifts ($^{13}C_{\alpha,i}$, $^{13}C_{\alpha,i-1}$ and optional $^{13}C_{\beta,i}$, $^{13}C_{\beta,i-1}$) associated with a particular amide nitrogen pair ($^{1}H^N_i$, $^{15}N_i$), unambiguous backbone amide H^N_i-H^N_j NOEs, and a semi-covalent structure of lowest U_{NOE} energy value. For comparison, also semi-covalent structure models derived from previous steps served as input structure. These data were additionally used in combination with residue type information resultant from backbone specific ^{15}N labeling and specific ^{13}C labeling at the carbonyl position C', which identify the backbone amide chemical shift of amino acids that succeeds in

primary sequence.

Pairwise runs of automated SA-MC backbone chemical shift assignments *with/without* a semi-covalent structure were performed using different extents of specific labeling patterns as input information as follows:

- As reference chemical shift assignments for a given chemical shift and backbone amide NOE database but resulting from the absence of any residue type labeling were generated.
- Initial, backbone specific ^{15}N residue type labeling was used for residues alanine, glycine, isoleucine, leucine, and valine.
- Then, for the same residue types the program input was supplemented using specific ^{13}C labeling for the carbonyl position.
- Afterwards, backbone specific ^{15}N labeling was employed for residue types phenylalanine, tyrosine, and tryptophane, whereas the former two were stated to be commutable to each other, i.e. one pair of backbone amide chemical shift were indicated to correspond either to phenylalanine or tyrosine residue type, respectively.
- Finally, in the same manner carbonyl specific ^{13}C labeling for these aromatic residue types were also taken into account.

Monte Carlo annealing parameters were set according ref [63] with exception of the number of swaps per temperature step, whose optimum value for a given molecular system was determined by empirical adjustment (i.e. by multiplying with an integer value until no significant changes regarding best results were achieved).

The benefit of the derived semi-covalent structures in respect to assignment quality were assessed by comparison of the assignment results obtained with or without it.

The analysis was further deepened by insertion of imperfections into the input chemical shift database of the α/β-fold protein cyclophilin D, whereas a semi-covalent structure model generated using experimental H^N-CH_3 NOEs only serves as input template for the chemical shift assignment.

These imperfections were introduced as follows:

- Random noise of $\pm$ 0.2 ppm were added to *each* intra $C_{\alpha,i}$ and inter $C_{\alpha,i-1}$ chemical shift entry (condition A).

- In addition, all inter $C_{\alpha,i-1}$ chemical shifts associated with H_i^N/N_i pairs of residues located outside of regular secondary structure were removed (condition B).
- At last, a selection of 25 intra $C_{\alpha,i}$ chemical shifts belonging to residues found in bend regions of the protein were also deleted (condition C).

For each input set of data, ten cycles of Monte Carlo simulated annealing were performed and ten assignment possibilities were calculated and analyzed with respect to correctness and uniqueness. A sequential assignment was considered correct, if it was reproduced in the majority of the cases and it was considered unique if it was consistently reproduced in ten out of ten cases.

3.5.2.5 Simulated annealing III

Once a assignment of the backbone chemical shifts is available, the using a *covalent* structure template, which enables application of chiral and torsion restraint for structure determination.

3.6 Automated binding site mapping using ROSA

The performance of the automated binding site mapping routine (section ?) was assessed using deposited data of a series of test protein-ligand complexes (table 3.20). The examples were selected according to the disposability of backbone $^1H^N$, ^{15}N, $^{13}C^\alpha$, and $^{13}C^\beta$ chemical shifts of both the *apo*- and *holo* form at the BMRB database as well as atom coordinates of a solved complex structure from the PDB repository (http://www.pdb.org). The types of ligands in the selection cover small peptides, cofactors, nucleotides and carbohydrates.

Table 3.20: Deposited protein-ligand complexes employed for testing automated binding site mapping.

Protein	**Ligand**	**BMRB entry**		**pdb code**	
CRBPII	*all-trans* Retinol	4681	4682	1B4M	1EII
Adenylate kinase	AP5A inhibitor	4152	4193	4AKE	1AKE
Ap4a hydrolase	ATP	4448	5054	1F3Y	1JKN
BtnCoA	biotin	4425	4426	3BDO	2BDO
Calmodulin	Rs20	5893	5896	-	1QTX
MBP	maltotriose	4986	4987	1JW4	3MBP
S100B	P53t	5895	4099	1QLK	1DT7
Snase	Thymidin	4052	4053	1JOR	1JOQ

The automated binding site mapping module in ROSA takes as input:

- the primary sequence of the protein
- A spin system list relating the backbone $^{1}H_i^N$ and $^{15}N_i$ chemical shifts of a residue i to its intra residue $^{13}C_i^{\alpha}$ resonance and to the $^{13}C_{i-1}^{\alpha}$ chemical shift of the preceding residue $i-1$. Optionally, the spin system for a unassigned residue i could further contain entries for $^{13}C_i^{'}/^{13}C_{i-1}^{'}$ and/or $^{13}C_i^{\beta}/^{13}C_{i-1}^{\beta}$ chemical shifts for a more unambiguous definition of a spin system fragment, if available.
- A file with sequential $^{1}H_i^N$-$^{1}H_i^N$ NOE contacts, if available. Since no NOESY restraint file was deposited for all test examples, a dummy file containing a senseless line was provided to make to program run. However, sequential $^{1}H_i^N$-$^{1}H_i^N$ NOE connectivities would alleviate the chemical shift degeneracy to establish unambiguous linking of the individual spinsystems.
- A list of spinsystem (working-)numbers, sorted according their $^{1}H_i^N$ and $^{15}N_i$ chemical shifts changes upon ligand binding. These changes were assessed by:

$$\Delta\delta_{apo\rightarrow holo} = \sqrt{\gamma_{^1H}/\gamma_{^{15}N} \cdot (^1H^N_{holo} - ^1 H^N_{apo})^2 + (^{15}N_{holo} - ^{15} N_{apo})^2} \qquad (3.2)$$

- A structure ($.pdb$) file containing the atom coordinates of the *apo* or *holo* form.
- A file ($.par$) including the parameters applied in the assignment part of the program.

A C_i^{α}-C_j^{α} distance cutoff of 8Å was used for evaluation of spatial neighboring residues (denoted as '*Distance cutoff value* ' at the PROSA input schedule see table ?). For good performance, at least two kinds of ^{13}C chemical shifts ($^{13}C^{\alpha}$, $^{13}C_i^{\beta}$, and/or $^{13}C_i^{'}$) were employed for chemical shift mapping on the primary sequence. As no significant improvement in the assignment performance was found by taken ^{13}C secondary chemical shifts into account, tabulated random coil for that purpose were used ('*Shift indexing*'). The constraint for chemical shift matching ('*Resolution*') was set to 0.2 ppm. Another parameter ('*Fraction of assignment possibilities*') defining a threshold with respect to the maximum score above which potential (chemical shift based) assignments are included for further scoring, as well as the cutoff above which solutions are accepted (given as the fraction of the score to the maximum) were leaved at their default values, 0.25 and 0.85, respectively.

Table 3.21: Description of adjustable parameters in ROSA.

Parameter	**Function**
Resolution	Chemical shift resolution in the cross peak file (0.1-0.3 ppm)
Shift indexing	Chemical shift indexing for sequential positioning of spinsystem fragments (random coil, calculated, distribution)
Fraction of assignment possibilities	determines the number of assignment possibilities for a given fragment
Distance cutoffvalue	denes the cutoff radius for defining the structural neighborhood of mapped spinsystem fragments
Neighbour reliability	threshold for accepting multiple solutions

The extent of the deposited assignments used for testing is given in table 3.22.

Table 3.22: Extent of available backbone chemical shifts.

Protein	**# of residues**	**# of Pro/Gly**	$^{13}\mathbf{C}_i^{\alpha}$/$^{13}\mathbf{C}_{i-1}^{\alpha}$ [%]	$^{13}\mathbf{C}_i^{\beta}$/$^{13}\mathbf{C}_{i-1}^{\beta}$ [%]
CRBPII	134	11/0	96.3/95.5	89.6/88.8
Adenylate kinase	214	9/19	99.0/98.5	93.5/93.0
Ap4a hydrolase	165	12/11	95.2/94.5	83.6/83.0
Biotinyl domain[a]	81	6/5	100.0/100.0	100.0/100.0
MBP	370	29/20	97.3/97.3	97.3/97.3
Snase	149	10/6	95.3/95.3	95.3/95.3

[a] of CH_3CO-CoA carboxylase

In the course of the binding site residues identification, the program lists the results as a table of spinsystems (each labeled with a working number) and a selection of possible sequential neighbors to support manual assignment of ambiguous positions. As final result a confrontation list with the provided shifted spinsystems introduced and the identified binding site residues is printed. As an additional feature of the program, a structure viewer window allows exploring the binding site interactively. It displays the sequential neighbors marked in the list or the tabulated positions, when marked in the structure view table.

Chapter 4

Results and Discussion

4.1 NMR analysis of folded proteins

This section summarizes the backbone and sidechain chemical shifts assignments, the structural results and some dynamics of the folded proteins *dog* ICln, *human* Cyclophilin D, and the LIM1 domain of *quail* CRP2. At all spectra no histidine cross peaks for $N^{\epsilon 2}/H^{N\epsilon 2}$ or $N^{\delta 1}/H^{N\delta 1}$ functions could be observed.

4.1.1 ICln

4.1.1.1 Resonance assignment

Sequential backbone $^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shift assignment of the entire wild type ICln sequence was obtained by manual resonance assignment of the truncated form (sample B, residue Met1 to residue Gln159, see figure 4.1) containing the folded part as well as manual chemical shift allocation of the construct (sample E) encompassing the intrinsic unfolded C-terminal part of the protein (residue Gln159 to residue His235).
Comparison of the $^{1}H^{N}$ and ^{15}N chemical shifts values observed at the $^{1}H^{N}$,^{15}N HSQC of (full length) wild type ICln (235 residues exclusive N-terminal 6xHis-tag) to their values found in the individual HSQC spectra of truncated and C-terminal ICln indicate a structural in-dependency of both part to each other (see figure 4.2). Remarkable differences in respect to thermodynamic ICln sample stabilities were observed depending on the presence of the unstructured C-terminal part. Samples preparations containing solely the N-terminal (folded) part of ICln (sample B and sample C) were found to be stable for about four weeks at room temperature. In contrast samples including the C-terminal part (sample A, sample D and sample E) were

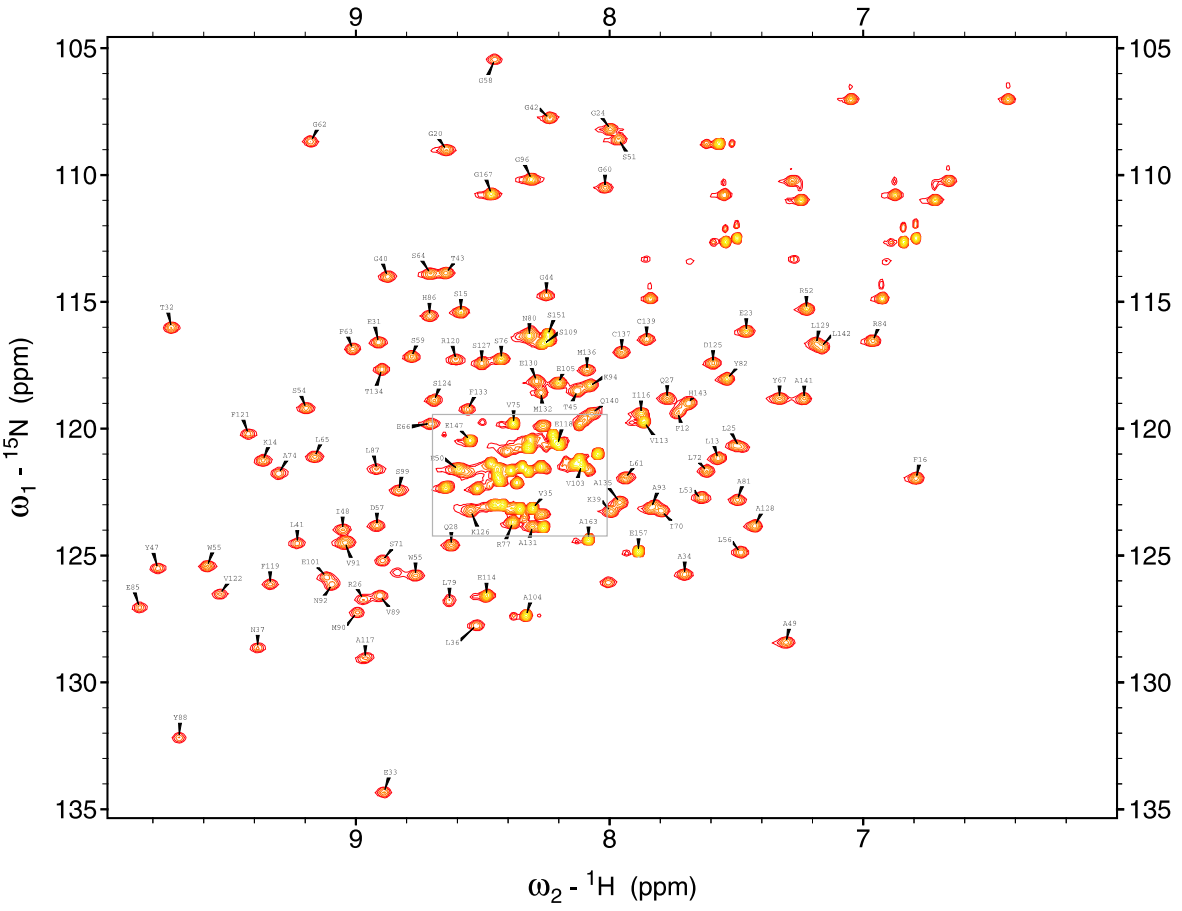


Figure 4.1: 800 MHz 2D $^{1}H^{N}$,^{15}N HSQC of N-terminal ICln containing the first 159 residues of the wild type sequence (sample B).

convenient for NMR measurement for only some days.

Furthermore, omitting the unstructured C-terminal part facilitated the sequential backbone assignment of amide resonances located around the middle of the (otherwise overcrowded) $^{1}H^{N}$,^{15}N HSQC. For that reasons, structure determination of the folded part of ICln was excursively performed by investigation of the truncated form of the protein (sample B). However, the presence of a 18 residue stretch (residue Glu89 to residue Ile107) found to be highly dynamical significantly complicates the whole assignment process. Because of the limited dynamic range of the analog to digital (ADC) converter peak highs of other residues in the triple resonances spectra employed appear to be perspicuously decreased relative to the very intense signals of that flexible part.

As visualized by an overlay of two $^{1}H^{N}$,^{15}N HSQC spectra related to a sample containing the whole N-terminal part (sample B) and an analog preparation (sample C), where the highly dynamical fragment residue Glu89 to residue Ile107 was replaced by Gly-Ser, respectively, indicate that the deletion of this flexible 18 residue part has no structural consequences for

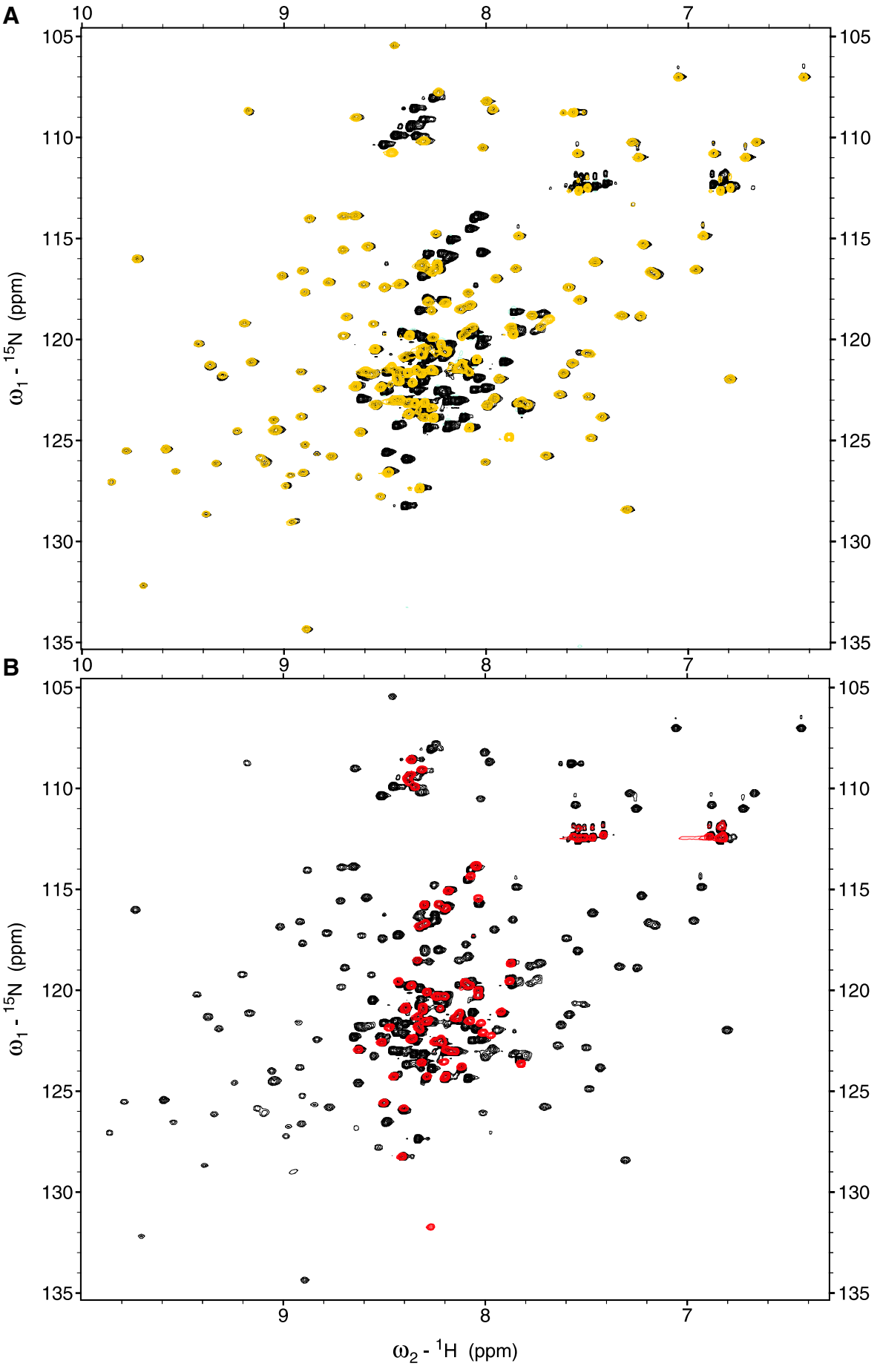


Figure 4.2: 2D $^{1}H^{N}$,^{15}N HSQC overlay of wild type ICln (black) and (A) its truncated form (blue, sample

the remaining structure of the protein but additionally resolve the degeneracy of the $^{1}H^{N}$,^{15}N chemical shift plane and lead to a remarkable gain of the signal intensities. Therefore (and because of limited sample availability) this preparation was used for titration experiments with the nucleoside AMP and two inositolphosphates as putative ligands. Because a uniformly *double* ^{13}C,^{15}N labeled preparation of sample B (including the flexible part) was still existent prior knowledge of this data, the benefit of sample C was not used for chemical shift and NOESY assignment of subsequent structure determination.

In total 183 of 223 non-proline $^{1}H^{N}$, ^{15}N backbone resonances were assigned (82.1%). $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and $^{13}C'$ chemical shift assignment was obtained for 202 residues. For 40 residues backbone amide $^{1}H^{N}$, ^{15}N assignment was not feasible either because they were not observable due to fast exchange with bulk water and/or rapid conformational change, or as result of strong overlap in the $^{1}H^{N}$,^{15}N chemical shift area typical for residues located in random coil structural parts (Flexible linker, N-terminal and C-terminal parts, see boxed regions in figure 4.1).

^{1}H and ^{13}C chemical shifts were allocated for 74.7% of aliphatic sidechains (exclusive starting methionine and 6xHis tag). In addition, ^{1}H,^{15}N chemical shift assignment for 5 (out of 7 observable) sidechain amide functions was possible (for 2 asparagine δ-NH_2 and for 3 ϵ-NH_2 glutamine groups). Furthermore, the $N^{\epsilon 1}/H^{N\epsilon 1}$ resonances of the only tryptophane residue (Trp55) could be unambiguously assigned.

The backbone chemical shift assignments of the truncated ICln form (sample B) have been deposited in the BioMagResBank under BMRB accession code number 5736, those of the intrinsic unstructured C-terminal part (sample E) are given in table 4.1. The complete obtained backbone and sidechain resonance assignment for ICln is given in table 4.2. As in the deposition, the sequential numbering for truncated ICln given here includes the N-terminal 6xHis tag connected by Glu-Leu to the starting methionine of the ICln wild type sequence (thus a integer of 9 have to be subtracted to relate the assignment number to the wild type sequence position).

Table 4.1: Backbone and sidechain chemical shift assignments for (N-terminal) truncated ICln (sample B, deposition numbering including Histag)

residue	$\delta(H^{N})$ [ppm]	$\delta(^{15}N)$ [ppm]	$\delta(^{13}C')$ [ppm]	$\delta(H^{\alpha}/C^{\alpha})$ [ppm]	others [ppm]
PHE12	7.77	119.40	174.97	4.76/47.31	3.38,3.05/39.40(H^{β}/C^{β}); 7.26(H^{δ}); 7.04(H^{ϵ}); ($^{\zeta}$)
LEU13	7.56	120.74	176.56	4.69/54.77	1.83,1.43/41.87(H^{β}/C^{β}); 1.03/26.07(H^{γ}/C^{γ}); 0.98/24.09($H^{\delta 1}/C^{\delta 1}$)
LYS14	9.39	121.16	175.11	4.79/54.89	2.05,1.91/36.83(H^{β}/C^{β});

residue	$\delta(H^N)$ [ppm]	$\delta(^{15}N)$ [ppm]	$\delta(^{13}C')$ [ppm]	$\delta(H^\alpha/C^\alpha)$ [ppm]	others [ppm]
					1.65,1.57/24.82(H^γ/C^γ); 1.79/29.33(H^δ/C^δ); 3.17,3.09/42.63(H^ϵ/C^ϵ)
SER15	8.60	115.22	174.52	5.03/57.68	3.73,3.73/64.03(H^β/C^β)
PHE16	6.82	121.84	-	4.91/55.07	3.15,3.00/37.48(H^β/C^β); 6.88(H^δ); 7.07(H^ϵ); -(ζ)
PRO19	-	-	175.37	4.14/62.73	2.03,1.55/32.50(H^β/C^β); 1.82,1.70/27.78(H^γ/C^γ); 3.08,2.13/(H^δ/C^δ)
GLY20	8.66	108.92	-	4.41,3.81/44.30	-
ALA22	-	-	178.41	4.25/-	1.43/18.10 (H^β/C^β)
GLU23 GLU23	7.49	116.19	176.46	4.37/57.12	2.36,1.98/29.80(H^β/C^β); 2.34,2.15/37.87(H^γ/C^γ)
GLY24	8.02	108.23	175.95	3.90,3.84/47.30	-
LEU25	7.60	121.13	176.01	3.88/57.34	1.64,1.18/42.03(H^β/C^β); 1.41/26.71(H^γ/C^γ); 0.97/24.57($H^{\delta1}/C^{\delta1}$); 1.0/25.93($H^{\delta2}/C^{\delta2}$)
ARG26	8.99	126.79	175.55	4.56/63.08	1.91,1.62/32.04(H^β/C^β); 1.73,1.62/27.68(H^γ/C^γ); 3.18,3.09/(H^δ/C^δ)
GLN27	7.76	118.79	172.62	4.44/56.14	32.14(C^β); 34.85(C^γ); 7.87,7.29/113.91(H^δ/N^δ)
GLN28	8.64	124.55	-	5.25/53.26	1.80,1.62/32.34(H^β/C^β); 2.00,1.88/33.02(H^γ/C^γ); 7.26,6.74/110.95(H^δ/N^δ)
GLN29	8.95	123.67	-	4.45/60.48	2.21/35.78(H^β/C^β); 7.07,6.45/106.98(H^δ/N^δ)
PRO30	-	-	174.36	4.78/-	2.31,2.03/33.57(H^β/C^β); 1.94,1.94/26.45(H^γ/C^γ); 3.71,3.63/(H^δ/C^δ)
GLU31	8.94	116.50	175.62	4.78/56.98	2.29,2.12/27.93(H^β/C^β); 2.23,2.23/37.00(H^γ/C^γ)
THR32	9.74	115.86	176.04	4.97/62.46	4.03/69.85(H^β/C^β); 1.22/21.89(H^γ/C^γ)
GLU33	8.92	134.31	175.67	5.34/56.01	2.06,2.06/31.41(H^β/C^β); 2.47,2.16/36.13(H^γ/C^γ)
ALA34	7.72	125.61	-	4.92/50.41	-0.12/19.44 (H^β/C^β)
VAL35	8.35	123.16	174.27	4.38/60.67	1.61/34.28(H^β/C^β); 0.45/19.72($H^{\gamma1}/C^{\gamma1}$); 0.61/20.69($H^{\gamma2}/C^{\gamma2}$)
LEU36	8.55	127.85	175.54	4.38/52.75	1.31,0.64/43.97(H^β/C^β); 0.90/27.39(H^γ/C^γ); 0.28/25.93($H^{\delta1}/C^{\delta1}$); -0.40/20.78($H^{\delta2}/C^{\delta2}$)
ASN37	9.40	128.79	177.94	4.39/53.96	2.95,2.77/36.91(H^β/C^β);

residue	$\delta(H^N)$ [ppm]	$\delta(^{15}N)$ [ppm]	$\delta(^{13}C')$ [ppm]	$\delta(H^\alpha/C^\alpha)$ [ppm]	others [ppm]
					7.6,6.91/113.19(H^δ/N^δ)
GLY38	8.40	102.40	173.51	4.19,3.43/45.56	-
LYS39	8.01	123.40	175.32	4.53/54.81	1.84,1.74/33.58(H^β/C^β); 1.38,1.32/24.61(H^γ/C^γ); 1.65,1.65/(H^δ/C^δ); 3.00/42.34(H^ϵ/C^ϵ)
GLY40	8.90	113.95	174.69	4.76,4.76/47.03	-
LEU41	9.27	124.55	176.23	4.57/54.54	1.76,1.50/43.15(H^β/C^β); 1.59/26.90(H^γ/C^γ); 0.84/22.82($H^{\delta1}/C^{\delta1}$); 0.60/26.42($H^{\delta2}/C^{\delta2}$)
GLY42	8.26	107.80	172.34	4.99,3.95/43.81	-
THR43	8.66	113.74	173.83	4.79/62.90	4.03/70.16(H^β/C^β); 1.24/21.76(H^γ/C^γ)
GLY44	8.26	114.59	171.26	4.12,3.32/45.88	-
THR45	8.15	118.62	171.67	4.78/60.90	3.53/70.66(H^β/C^β); 0.18/20.69(H^γ/C^γ)
LEU46	8.03	126.04	173.79	4.79/53.16	1.95,1.10/45.85(H^β/C^β); 1.24/27.19(H^γ/C^γ); 0.48/27.19($H^{\delta1}/C^{\delta1}$); 0.2/22.82($H^{\delta2}/C^{\delta2}$)
TYR47	9.80	125.40	175.19	4.73/57.48	2.93,2.55/39.44(H^β/C^β); 6.84(H^δ); 6.57(H^ϵ)
ILE48	9.04	123.92	173.75	4.76/61.64	1.89/37.61(H^β/C^β); 0.86/17.87($H^{\gamma2}/C^{\gamma2}$); 0.84/14.57(H^δ/C^δ)
ALA49	7.28	128.37	176.16	3.90/49.19	1.57/21.77 (H^β/C^β)
GLU50	8.60	121.74	177.72	3.87/60.40	2.19,2.09/29.53(H^β/C^β); 2.27,2.27/37.19(H^γ/C^γ)
SER51	8.00	108.66	174.91	4.77/60.68	4.03,3.87/63.26(H^β/C^β)
ARG52	7.24	115.34	172.82	4.11/55.16	2.12,1.43/30.30(H^β/C^β); 1.31,1.01/24.58(H^γ/C^γ); 2.97,2.71/(H^δ/C^δ)
LEU53	7.64	122.64	176.37	5.01/54.12	1.39,1.34/44.29(H^β/C^β); 1.36/27.00(H^γ/C^γ); 0.72/26.61($H^{\delta1}/C^{\delta1}$); 0.55/25.35($H^{\delta2}/C^{\delta2}$)
SER54	9.21	119.04	170.43	5.43/58.14	3.84,3.54/67.11(H^β/C^β)
TRP55	8.78	125.71	174.27	5.45/56.69	2.25,2.41/32.47(H^β/C^β); 6.55 (H^δ); 9.61/125.41($H^{\epsilon1}/N^{\epsilon1}$);
LEU56	7.50	124.77	175.55	4.42/53.56	1.49,1.10/48.56(H^β/C^β); 1.09/26.80(H^γ/C^γ); 0.38/24.77($H^{\delta1}/C^{\delta1}$); -0.12/23.50($H^{\delta2}/C^{\delta2}$)

residue	$\delta(H^N)$ [ppm]	$\delta(^{15}N)$ [ppm]	$\delta(^{13}C')$ [ppm]	$\delta(H^\alpha/C^\alpha)$ [ppm]	others [ppm]
ASP57	8.95	123.82	177.91	4.06/52.48	3.22,2.74/41.05(H^β/C^β)
GLY58	8.47	105.45	175.22	4.15,3.18/46.76	-
SER59	8.80	117.13	174.43	4.61/58.06	4.03,3.92/64.44(H^β/C^β)
GLY60	8.02	110.44	172.54	4.05,3.41/46.03	-
LEU61	7.96	122.01	176.77	4.59/53.06	1.72,1.33/44.36(H^β/C^β); 1.41/26.90(H^γ/C^γ); 0.81/24.86($H^{\delta1}/C^{\delta1}$); 0.72/24.18($H^{\delta2}/C^{\delta2}$)
GLY62	9.19	108.61	172.30	4.60,4.60/47.36	-
PHE63	9.03	116.71	172.43	5.39/56.34	3.34,3.34/42.03(H^β/C^β); 6.98(H^δ); 6.98(H^ϵ); -(ζ)
SER64	8.74	113.74	172.36	5.44/56.87	3.37,3.37/66.62(H^β/C^β)
LEU65	9.16	121.16	175.30	4.79/52.72	1.47,0.94/44.87(H^β/C^β); 1.38/26.61(H^γ/C^γ); 0.64/23.60($H^{\delta1}/C^{\delta1}$); 0.30/24.77($H^{\delta2}/C^{\delta2}$)
GLU66	8.72	119.89	174.50	4.37/54.84	2.16,1.88/30.64(H^β/C^β); 2.37,2.18/35.84(H^γ/C^γ)
TYR67	7.34	118.83	-	4.57/53.57	2.85,2.61/40.74(H^β/C^β); 7.35(H^δ); 7.56(H^ϵ)
PRO68	-	-	176.98	-	37.61(C^β);
THR69	7.56	104.50	173.82	4.77/62.13	4.49/70.01(H^β/C^β); 1.26/22.36(H^γ/C^γ)
ILE70	7.82	123.29	175.73	4.36/62.25	2.11/38.35(H^β/C^β); 1.03,1.03/23.21($H^{\gamma1}/C^{\gamma1}$); 0.94/17.35($H^{\gamma2}/C^{\gamma2}$); 0.66/14.28(H^δ/C^δ)
SER71	8.92	125.19	174.79	4.52/59.17	3.86,3.81/64.34(H^β/C^β)
LEU72	7.64	121.58	-	4.72/56.06	1.90,1.63/46.06(H^β/C^β); 1.62/27.58(H^γ/C^γ); 1.05/24.97($H^{\delta1}/C^{\delta1}$); 0.9/25.47($H^{\delta2}/C^{\delta2}$)
HIS73	-	-	173.53	6.00/55.33	3.20,3.11/34.96(H^β/C^β)
ALA74	9.34	121.80	177.73	4.57/52.31	1.33/22.93 (H^β/C^β)
VAL75	8.39	119.71	176.57	4.77/56.87	-/30.47(H^β/C^β);
SER76	8.44	117.16	-	4.62/58.46	3.87,3.87/63.70(H^β/C^β)
ARG77	8.42	123.92	-	4.36/56.68	1.87,1.82/32.93(H^β/C^β); 1.43,1.43/24.77(H^γ/C^γ); 2.98,2.98/(H^δ/C^δ)
ASP78	-	-	177.00	4.61/-	2.85,2.62/40.73(H^β/C^β)
LEU79	8.65	126.79	176.14	4.80/55.58	1.71,1.48/39.82(H^β/C^β); 1.74/26.80(H^γ/C^γ); 0.89/25.25($H^{\delta1}/C^{\delta1}$); 0.72/21.56($H^{\delta2}/C^{\delta2}$)
ASN80	8.33	116.29	176.14	4.47/54.91	2.85,2.71/38.68(H^β/C^β); 7.8,6.94/114.82(H^δ/N^δ)

residue	$\delta(H^N)$ [ppm]	$\delta(^{15}N)$ [ppm]	$\delta(^{13}C')$ [ppm]	$\delta(H^\alpha/C^\alpha)$ [ppm]	others [ppm]
ALA81	7.53	122.64	177.51	4.20/53.65	1.41/19.44 (H^β/C^β)
TYR82	7.54	117.98	-	4.55/57.17	2.97,2.71/41.71(H^β/C^β); 6.97(H^δ); 6.55(H^ϵ)
PRO83	-	-	175.47	-	32.00(C^β); -,-/-(H^γ/C^γ); -,-/(H^δ/C^δ)
ARG84	6.99	116.50	175.78	4.74/54.61	2.03,1.59/35.91(H^β/C^β); 1.68,1.62/27.10(H^γ/C^γ); 3.30,3.26/(H^δ/C^δ)
GLU85	9.88	127.11	175.95	4.74/58.34	2.03,1.99/29.29(H^β/C^β); 2.44,2.35/37.68(H^γ/C^γ)
HIS86	8.73	115.44	171.05	4.80/56.42	3.20,2.99/32.12(H^β/C^β)
LEU87	8.92	121.70	174.17	5.14/54.06	2.23,1.46/45.18(H^β/C^β); 1.14/24.18($H^{\delta1}/C^{\delta1}$);
TYR88	9.71	132.19	174.03	5.37/56.94	3.19,2.94/40.94(H^β/C^β); 7.07(H^δ); 6.59(H^ϵ)
VAL89	8.92	126.67	173.02	4.72/61.51	2.05/36.41(H^β/C^β); 1.05/21.85($H^{\gamma1}/C^{\gamma1}$); 0.90/22.82($H^{\gamma2}/C^{\gamma2}$)
MET90	9.00	127.32	173.91	5.13/53.72	2.44,2.29/32.73(H^β/C^β); 2.09,2.09/35.35(H^γ/C^γ); -/-(H^ϵ/C^ϵ)
VAL91	9.06	124.38	175.06	4.97/60.56	2.22/35.95(H^β/C^β);
ASN92	9.10	126.04	173.82	5.19/52.47	2.88,2.77/38.22(H^β/C^β); 7.5,6.89/110.75(H^δ/N^δ)
ALA93	7.87	123.08	174.90	4.32/52.11	1.22/21.26 (H^β/C^β)
LYS94	8.10	118.40	-	4.32/55.95	1.72,1.58/33.53(H^β/C^β); 1.34,1.27/25.06(H^γ/C^γ); 1.62,1.62/(H^δ/C^δ); 2.92,2.92/42.05(H^ϵ/C^ϵ)
PHE95	-	-	175.57	4.49/57.85	2.61,2.61/40.88(H^β/C^β); 6.95(H^δ); 6.95(H^ϵ); -(ζ)
GLY96	8.33	110.14	-	4.77,3.88/45.34	-
GLU98	-	-	175.47	-	32.15(C^β);
SER99	8.84	122.43	-	4.77/56.68	3.08,2.61/63.70(H^β/C^β)
LYS100	-	-	178.75	-	30.63(C^β)
GLU101	9.14	125.83	-	4.16/64.60	3.11,2.58/35.85(H^β/C^β)
SER102	-	-	174.45	-	63.91(C^β)
VAL103	8.13	121.55	175.86	4.19/62.03	2.12/32.82(H^β/C^β); 0.91/21.27($H^{\gamma1}/C^{\gamma1}$); 0.93/20.30($H^{\gamma2}/C^{\gamma2}$)
ALA104	8.36	127.42	177.66	4.20/52.76	1.39/19.21 (H^β/C^β)
GLU105	8.24	118.09	-	-	-
ASP112	-	-	-	-	-
VAL113	7.88	119.68	175.99	4.11/62.04	2.05/33.04(H^β/C^β); 0.89/21.37($H^{\gamma1}/C^{\gamma1}$);

residue	$\delta(H^N)$ [ppm]	$\delta(^{15}N)$ [ppm]	$\delta(^{13}C')$ [ppm]	$\delta(H^\alpha/C^\alpha)$ [ppm]	others [ppm]
					0.88/20.59($H^{\gamma 2}/C^{\gamma 2}$)
GLU114	8.49	126.58	-	4.54/54.35	2.00,1.88/29.98(H^β/C^β); 2.28,2.28/35.83(H^γ/C^γ)
ILE116	7.89	119.25	175.10	4.71/60.13	1.79/41.09(H^β/C^β); 1.45,1.13/27.39($H^{\gamma 1}/C^{\gamma 1}$); 0.86/17.87($H^{\gamma 2}/C^{\gamma 2}$); 0.83/13.31($H^\delta/C^\delta$)
ALA117	8.96	129.01	174.81	4.72/51.06	1.25/21.68 (H^β/C^β)
GLU118	-	-	174.78	4.72/55.15	2.03,1.84/32.68(H^β/C^β); 2.22,1.91/37.68(H^γ/C^γ)
PHE119	9.37	126.04	174.99	5.23/56.65	2.79,1.95/42.24(H^β/C^β); 6.58(H^δ); 6.80(H^ϵ)
ARG120	8.63	117.13	174.51	5.16/54.19	32.18(C^β);
PHE121	9.45	120.31	175.56	5.64/56.41	2.76,3.00/40.62(H^β/C^β); 7.00(H^δ); 7.23(H^ϵ)
VAL122	9.55	126.58	-	4.78/58.89	2.52/32.91(H^β/C^β); 1.10/22.34($H^{\gamma 1}/C^{\gamma 1}$); 0.69/18.82($H^{\gamma 2}/C^{\gamma 2}$)
PRO123	-	-	176.34	4.44	2.08,1.84/31.65(H^β/C^β)
SER124	8.71	118.83	174.90	4.34/63.31	4.17,4.09/61.11(H^β/C^β)
ASP125	7.61	117.56	175.20	4.92/51.87	2.88,2.58/40.30(H^β/C^β)
LYS126	8.56	123.29	178.75	4.07/58.29	2.05,2.01/31.78(H^β/C^β); 1.70,1.49/23.99(H^γ/C^γ); 1.77,1.70/(H^δ/C^δ);
LYS126	8.56	123.29	178.75	4.07/58.29	3.16/42.56(H^ϵ/C^ϵ)
SER127	8.52	117.13	175.16	4.78/60.95	4.08,4.08/63.12(H^β/C^β)
ALA128	7.45	123.70	177.78	4.41/53.09	1.44/19.64 (H^β/C^β)
LEU129	7.20	116.71	178.15	4.19/58.95	1.92,1.37/41.87(H^β/C^β); 1.70/28.07(H^γ/C^γ); 0.96/26.03($H^{\delta 1}/C^{\delta 1}$); 0.95/25.26($H^{\delta 2}/C^{\delta 2}$)
GLU130	8.32	118.09	178.92	4.11/60.88	2.11,2.11/28.55(H^β/C^β); 2.40,2.32/37.00(H^γ/C^γ)
ALA131	8.32	123.82	181.39	4.21/55.39	1.53/18.07 (H^β/C^β)
MET132	8.31	118.62	176.91	3.71/59.43	2.38,2.04/35.64(H^β/C^β); 2.75,2.34/31.37(H^γ/C^γ); 0.84/14.67(H^ϵ/C^ϵ)
PHE133	8.57	119.25	177.81	4.06/62.36	3.30,3.08/39.84(H^β/C^β); 7.21(H^δ); -(H^ϵ); -($^\zeta$)
THR134	8.92	117.56	176.08	3.72/67.54	4.23/68.87(H^β/C^β); 1.29/21.82(H^γ/C^γ)
ALA135	7.98	122.97	179.45	4.19/55.40	1.38/18.46 (H^β/C^β)
MET136	8.11	117.56	177.95	3.69/59.81	1.60,2.02/31.95(H^β/C^β); 2.46,2.13/33.31(H^γ/C^γ); 0.88/17.97(H^ϵ/C^ϵ)
CYS137	7.96	116.92	-	4.06/63.44	3.09,2.77/26.51(H^β/C^β)

residue	$\delta(H^N)$ [ppm]	$\delta(^{15}N)$ [ppm]	$\delta(^{13}C')$ [ppm]	$\delta(H^\alpha/C^\alpha)$ [ppm]	others [ppm]
GLU138	-	-	179.26	3.89/59.36	2.14,1.93/29.72(H^β/C^β); 2.51,2.15/36.71(H^γ/C^γ)
CYS139	7.86	116.50	175.96	4.09/63.92	2.03,1.86/28.60(H^β/C^β)
GLN140	8.09	119.47	177.82	3.41/59.05	2.04,1.87/28.71(H^β/C^β); 2.25,2.25/34.18(H^γ/C^γ); 7.30,6.67/110.24(H^δ/N^δ)
ALA141	7.24	118.83	177.81	4.14/53.33	1.42/18.45 (H^β/C^β)
LEU142	7.19	116.71	176.01	4.16/55.13	1.74,1.40/41.79(H^β/C^β); 1.81/26.22(H^γ/C^γ); 0.75/22.92($H^{\delta1}/C^{\delta1}$); 0.76/22.14($H^{\delta2}/C^{\delta2}$)
HIS143	7.74	118.83	-	4.66/53.64	3.16,3.18/30.13(H^β/C^β)
PRO146	-	-	177.19	4.44/-	2.33,2.33/32.33(H^β/C^β); 2.00,2.00/27.50(H^γ/C^γ); 3.86,3.86/(H^δ/C^δ)
GLU147	8.56	120.52	-	4.27/56.68	1.95/30.43(H^β/C^β); 2.31,2.31/36.30(H^γ/C^γ)
ASP155	-	-	176.31	-	41.49(C^β)
GLY156	7.59	108.65	173.35	4.77,3.90/45.51	-
GLU157	7.90	124.88	-	4.19/57.14	2.29,2.14/30.47(H^β/C^β); 1.92/36.50(H^γ/C^γ)
GLU162	-	-	176.47	-	30.19(C^β)
ALA163	8.10	124.34	-	4.22/52.93	1.33/19.18 (H^β/C^β)
GLN166	-	-	176.54	-	29.55(C^β)

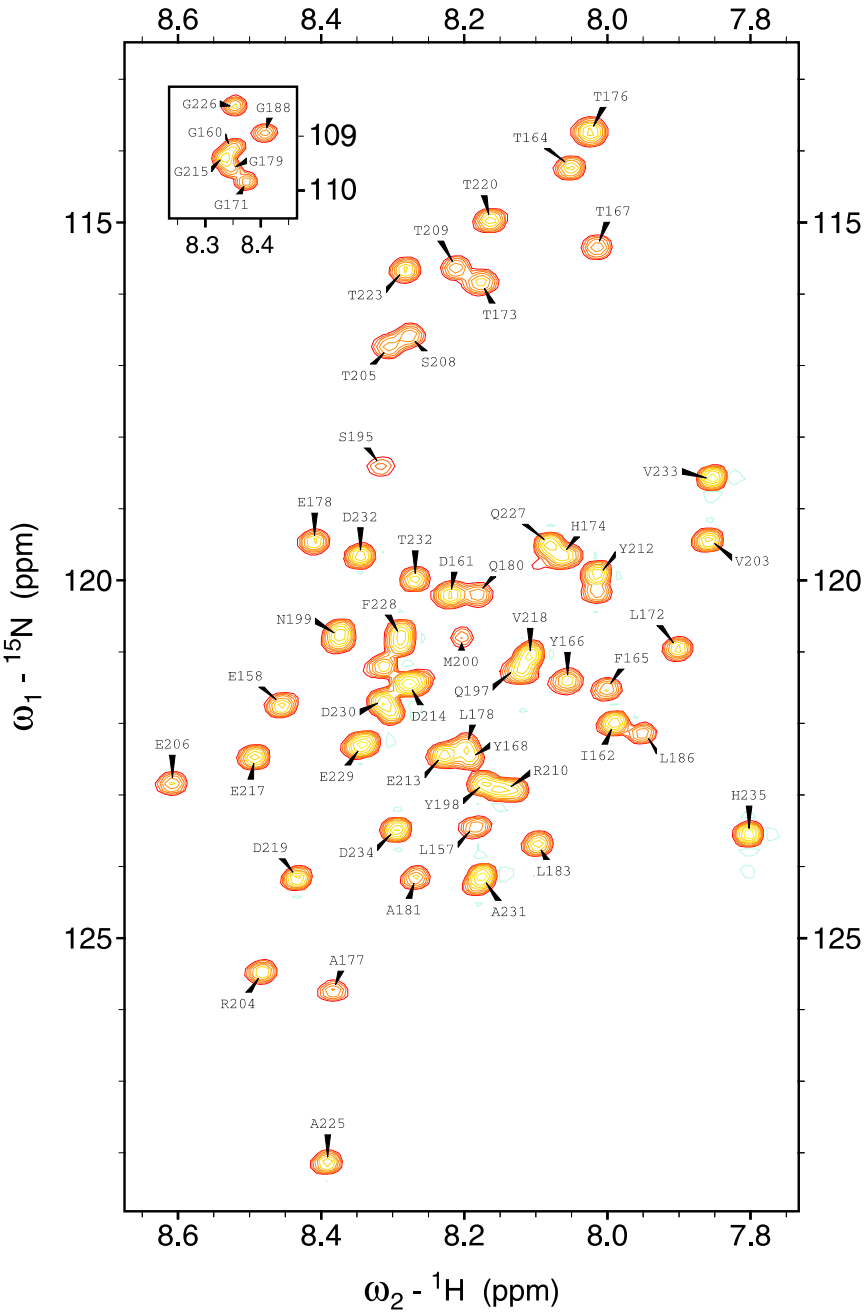


Figure 4.3: 2D $^{1}H^{N}$,^{15}N HSQC of the C-terminal ICln (sample E). Glycine residue are shown in the box (left top).

Table 4.2: Backbone chemical shifts for C-terminal ICln (sample E, wild type sequence numbering).

Residue	δ(**H**^{N}) [ppm]	δ(**^{15}N**) [ppm]	δ(**^{13}C'**) [ppm]	δ(**C**$_{\alpha}$) [ppm]	δ(**C**$_{\beta}$) [ppm]
GLN159	8.35	118.49	176.45	57.53	30.12
GLY160	8.39	109.39	173.69	45.65	-
ASP161	8.26	120.34	175.89	54.66	41.67
ILE162	8.03	121.96	174.60	58.84	38.65
PRO163	-	-	176.74	-	32.46
THR164	8.09	114.33	174.04	62.40	70.34
PHE165	8.04	121.73	174.98	57.71	40.17
TYR166	8.10	121.50	175.60	58.15	39.64
THR167	8.05	115.48	173.98	61.83	70.58
TYR168	8.23	122.43	175.95	58.83	39.05
GLU169	8.37	122.43	176.59	57.16	30.28
GLU170	-	-	177.40	-	30.15
GLY171	8.39	109.93	174.59	45.42	-
LEU172	7.94	121.00	177.83	55.84	42.72
SER173	8.22	115.72	174.66	58.94	63.58
HIS174	8.09	119.65	176.74	56.42	32.85
LEU175	8.24	122.43	177.70	55.74	42.81
THR176	8.06	113.86	174.64	61.94	70.72
ALA177	8.42	125.90	178.47	53.63	19.32
GLU178	8.45	119.65	177.65	57.84	30.22
GLY179	8.36	110.16	174.71	46.14	-
GLN180	8.21	120.34	176.30	56.81	29.64
ALA181	8.31	124.28	178.51	53.59	19.27
THR182	8.10	113.86	174.88	63.18	70.04
LEU183	8.14	123.81	177.82	56.20	42.18
ARG185	-	-	175.20	-	30.80
LEU186	7.99	122.16	177.68	55.35	42.49
GLU187	-	-	177.56	-	30.40
GLY188	8.33	108.77	174.63	45.63	-
VAL194	-	-	176.45	-	33.08
SER195	8.36	118.49	174.97	58.70	63.67
SER196	-	-	177.10	-	64.21
GLN197	8.16	121.27	176.77	57.04	30.97
TYR198	8.21	122.89	-	56.54	-
ASN199	-	-	175.23	-	39.19
MET200	8.24	120.81	176.09	56.17	32.62
ALA201	8.22	124.28	178.20	53.18	19.43
GLY202	8.29	108.08	174.00	45.69	-
VAL203	7.90	119.42	176.09	62.63	33.25
ARG204	8.52	125.43	176.51	56.33	31.31
THR205	8.34	116.64	174.85	62.43	70.28
GLU206	8.65	122.89	176.22	56.94	30.19
ASP207	-	-	176.51	-	41.63

Residue	δ(**H**^{N}) [ppm]	δ(^{15}**N**) [ppm]	δ(^{13}**C'**) [ppm]	δ(**C**$_\alpha$) [ppm]	δ(**C**$_\beta$) [ppm]
SER208	8.31	116.64	175.13	59.12	64.22
THR209	8.25	115.72	174.65	62.48	69.49
ARG210	8.17	122.93	175.43	56.06	30.98
ASP211	-	-	175.81	-	41.52
TYR212	8.05	119.88	175.80	58.31	39.29
GLU213	8.27	122.43	176.02	56.49	30.60
ASP214	8.32	121.38	176.91	55.14	41.68
GLY215	8.40	109.47	174.43	45.50	-
MET216	8.12	119.65	176.18	55.86	33.01
GLU217	8.53	122.43	176.35	56.98	30.45
VAL218	8.15	121.04	175.70	62.44	33.53
ASP219	8.47	124.28	176.49	54.60	41.61
THR220	8.20	115.02	174.69	61.86	70.10
THR221	8.31	120.11	172.92	60.71	69.57
PRO222	-	-	177.06	-	32.53
THR223	8.32	115.72	174.61	62.60	70.24
VAL224	8.18	123.12	175.72	62.40	33.32
ALA225	8.43	128.21	178.20	53.05	19.58
GLY226	8.38	108.31	173.94	45.87	-
GLN227	8.11	119.65	176.67	56.54	29.87
PHE228	8.33	120.81	175.71	58.03	39.78
GLU229	8.38	122.43	175.93	56.99	30.84
ASP230	8.35	121.73	175.95	54.68	41.57
ALA231	8.21	124.16	177.48	53.00	19.87
ASP232	8.38	119.65	176.22	54.66	41.44
VAL233	7.89	118.49	175.77	62.28	33.37
ASP234	8.33	123.58	175.16	54.86	41.57

Using the more conformationally sensitive (and assigned) $^{13}C^{\alpha}$ and $^{13}C^{\alpha}$ resonances, the consensus chemical shift index (CSI) was employed [194] to visualize the secondary structure elements of ICln. The obtained $\Delta^{13}C^{\alpha}$-$\Delta^{13}C^{\beta}$ secondary chemical shift values for the N-terminus of ICln (residue Met1 to residue Gln159) encompassing the structural parts of the protein indicate seven β strands and one α-helix (see figure 4.4.A). The exact termination points of these secondary elements were verified by dint of sequential $^{1}H^{N}$-$^{1}H^{N}$ and $^{1}H^{N}$-$^{1}H^{\alpha}$ NOEs, whose patterns analysis generally revealed a good agreement with the secondary structure prediction obtained by the CSI.

Furthermore, some preference for a α-helical structure was found for residue Gln180 to residue Glu184 of the remaining intrinsic unfolded C-terminal part (residue Gln159 to residue His235, figure 4.4.B).

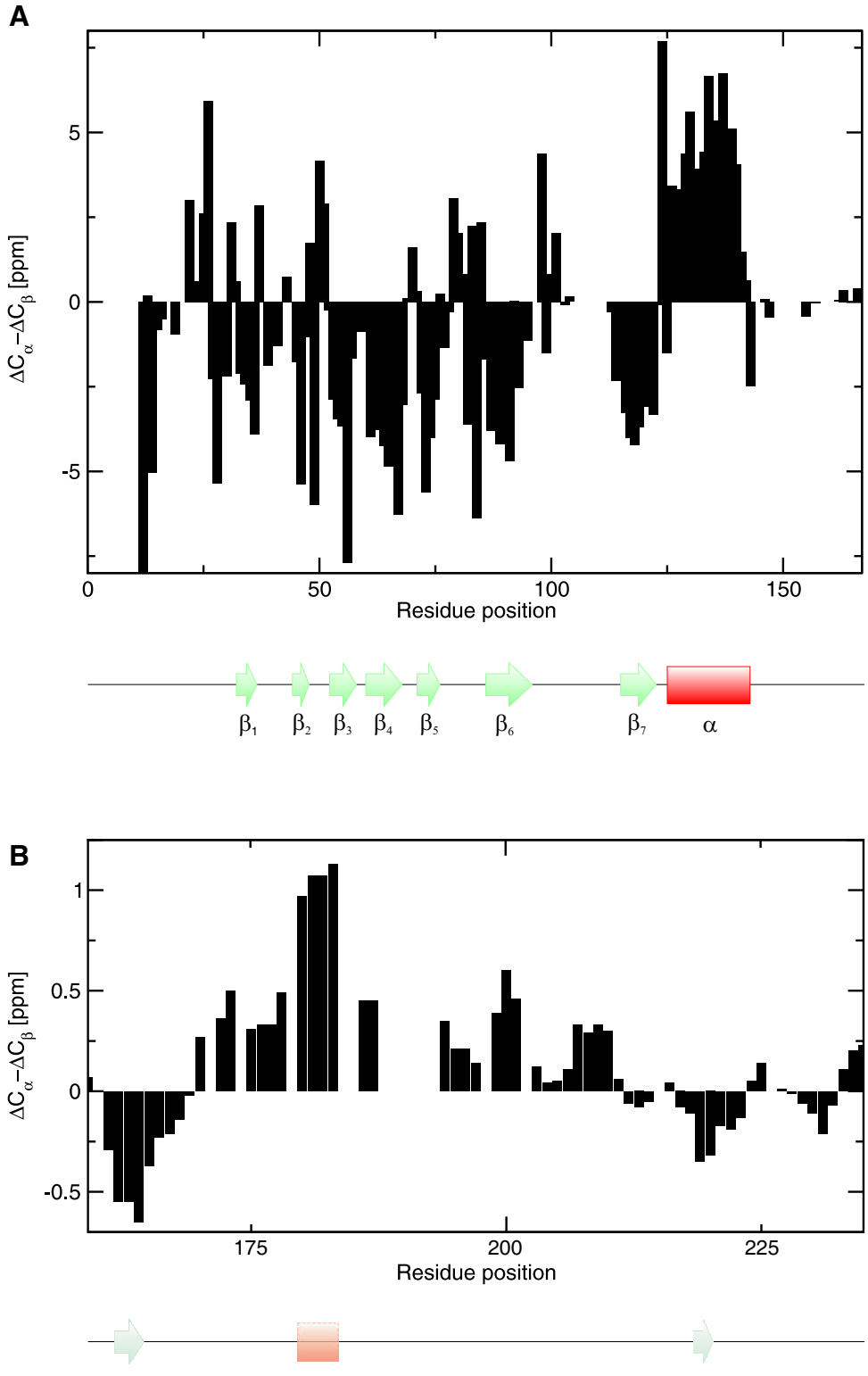


Figure 4.4: $\Delta^{13}C^{\alpha}$-$\Delta^{13}C^{\beta}$ secondary chemical shifts versus residue position for (A) truncated (N-terminal) and (B) C-terminal ICln.

4.1.1.2 Stereo assignments and χ_1 angle restraints

Via ^{13}CO-spin echo constant time HSQC experiments $J_{C^\gamma C'}$ and $^3J_{C^\gamma N}$ couplings between the backbone carbonyl ^{13}C' and amide nitrogen ^{15}N, and the methyl carbons -$^{13}CH_3$ of valine, isoleucine ($C^{\gamma 2}$), and threonines, respectively, were measured (table 4.3). From this data and equation 2.7, the rotameric states (g^+, g^-, or t) of the methyl carbons C$^\gamma$ to the intra-residue backbone carbonyl for residues Val35, Thr69, Ile70, Val89, Val103, Ile116, and Val122 were inferred. In addition, $J_{C^\beta C'}$ coupling values for residue type alanine were obtained.

Table 4.3: Determined $J_{C^\gamma C'}$ and $^3J_{C^\gamma N}$ couplings and associated uncertainties for truncated ICln (sample B)

Residue	$^{2,3}J_{C^{\beta,\gamma}C'}$	Error	$^3J_{C^\gamma N}$	Error
Val35($C^{\gamma 1}$)	3.5	0.2	0.1<	0.1
Val35($C^{\gamma 2}$)	1.13	0.4	1.4	0.2
Ala49(C^β)	0.3	0.1	-	-
Thr69($C^{\gamma 2}$)	2.5	0.3	1.3	0.4
Ile70($C^{\gamma 2}$)	0.1<	0.1	1.9	0.2
Val89($C^{\gamma 1}$)	3.4	0.2	0.1<	0.1
Val89($C^{\gamma 2}$)	1.2	0.4	2.3	0.4
Ala93(C^β)	1.3	0.4	-	-
Val103($C^{\gamma 1}$)	2.0	0.3	0.5	0.3
Val103($C^{\gamma 2}$)	2.3	0.3	0.1<	0.1
Ile116($C^{\gamma 2}$)	1.7	0.4	0.8	0.4
Ala117(C^β)	1.0	0.4	-	-
Val122($C^{\gamma 1}$)	0.9	0.4	1.3	0.1
Val122($C^{\gamma 2}$)	3.0	0.3	0.1<	0.1
Ala128(C^β)	2.9	0.3	-	-
Ala131(C^β)	3.0	0.2	-	-
Ala135(C^β)	3.7	0.1	-	-
Ala141(C^β)	2.6	0.3	-	-

4.1.1.3 ^{15}N-Relaxation

T_1, and T_2 values for sample B could be extracted for 102 (out of 110) non-proline assigned backbone nitrogen atoms measured at spectrometer frequencies of 500 and 800 MHz, respectively. Resonance intensities influenced by other signal positions (mainly arising from the high dynamical loops β_{56} and β_{67}) were rejected from the analysis. Analog relaxation data was acquired for sample E using all assigned backbone ^{15}N resonance positions.

As shown in figure 4.5.A the T_1 values of of truncated ICln (sample B) and the unfolded C-terminus (sample E) are essentially uniform along the primary sequence except for the loops

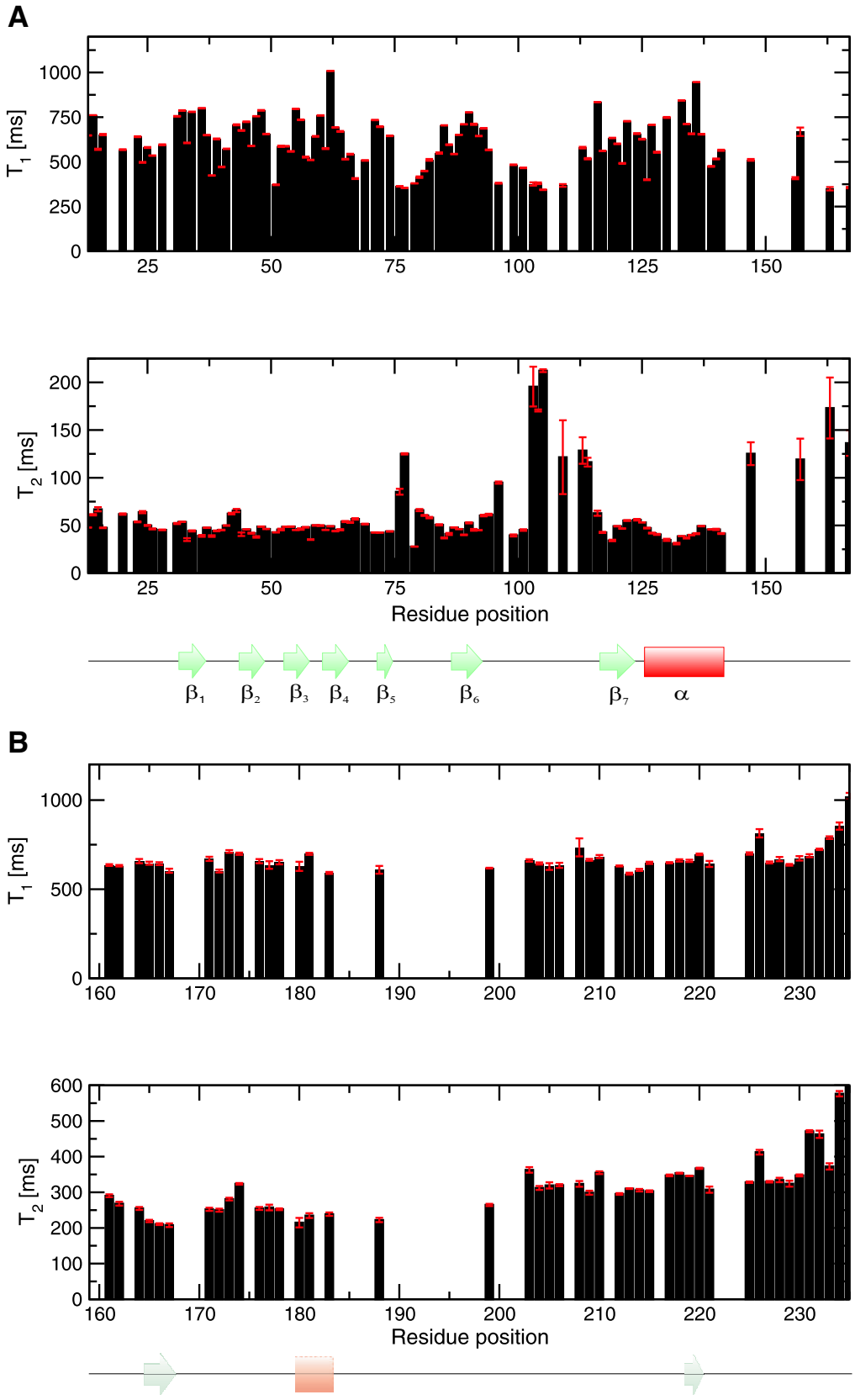


Figure 4.5: ^{15}N-Relaxation T_1 and T_2 values obtained at 800MHz and 298K of (A) N-terminal and (B) C-terminal ICln. For the N-terminal part (containing the folded segments) location of the secondary structure elements as inferred from $\Delta^{13}C^{\alpha}$-$\Delta^{13}C^{\beta}$ secondary chemical shifts and NOE pattern analysis are given at the bottom. In (B) preferences for α-helical and elongated structure elements as indicated from $\Delta^{13}C^{\alpha}$-$\Delta^{13}C^{\beta}$ secondary chemical shift values are drawn in transparent.

β_5/β_6 and β_6/β_7, and the C-terminal residues as a result of increased conformational flexibility. Beside for that regions significant increased T_2 values were observed for the unstructured C-terminal part (residue Leu133 to residue His235; wild type sequence numbering without His-tag, see figure 4.5.B), whereas sequence regions of some secondary structure preference (extended conformation for residue Ile162 to residue The165, and β-branched residues Thr219/Thr220, as well as a α-helical preference for residue Ala180 to residue Leu183) show in average slightly more decreased T_2 values compared to the residual intrinsic unfolded C-termini (230ms likened to 260ms, respectively).

4.1.1.4 ICln folds in a Pleckstrine homology-like structure

ICln exhibits a pleckstrin homology (PH) domain topology. It consists of a pair of nearly orthogonal anti-parallel β-sheets that form a collapsed β-barrel. The two sheets are linked by the continuation of β-strands β_1 and β_2 and a linker loop β_4/β_5. Near the N-Terminus this barrel like structure is capped on one side by a amphipathic C-terminal α-helix. The amino and carboxy termini are located at the same side of the protein, but protrude in opposite directions (see also figure 4.6.A).

The first β-sheet consists of three anti-parallel β-strands comprising strand β_2 (residue Gly44 to residue Ala49) and strand β_3 (residue Leu53 to residue Asp57), which are connected by a β-hair pin type I turn, and strand β_4 (residue Gly62 to residue Ser64). A seven residue (residue Leu65 to residue Ser71) low-dynamic linker connects both β-sheets between strand β_4 and strand β_5 on the bottom of the β-barrel at the N- and C-terminal side.

The second β-sheet contains strand β_1 (residue Glu33 to residue Val35), strand β_5 (residue Leu72 to residue Ala74), strand β_6 (residue His86 to residue Asn92), and strand β_7 (residue Ile116 to residue Val122). As mentioned above, ^{15}N relaxation data (figure ?) reveal a high dynamical behavior of the segments connecting strand β_5 to strand β_6 (loop $\beta_{5/6}$), and strand β_6 to strand β_7 (loop $\beta_{6/7}$), respectively. All assigned residues in that parts show low deviations from the random coil values for the backbone $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts indicating a complete absence of any secondary structure. Furthermore, all observable amide protons in both loops show strong NOE cross correlations with the water resonance in the ^{15}N-edited NOESY spectrum, suggesting their fast exchange with solvent.

A conspicuously feature of the high dynamical 23 residue $\beta_{6/7}$ linker loop is its accommodation of a remarkable number of acidic residues (seven glutamates: Glu97, Glu98, Glu101, Glu105,

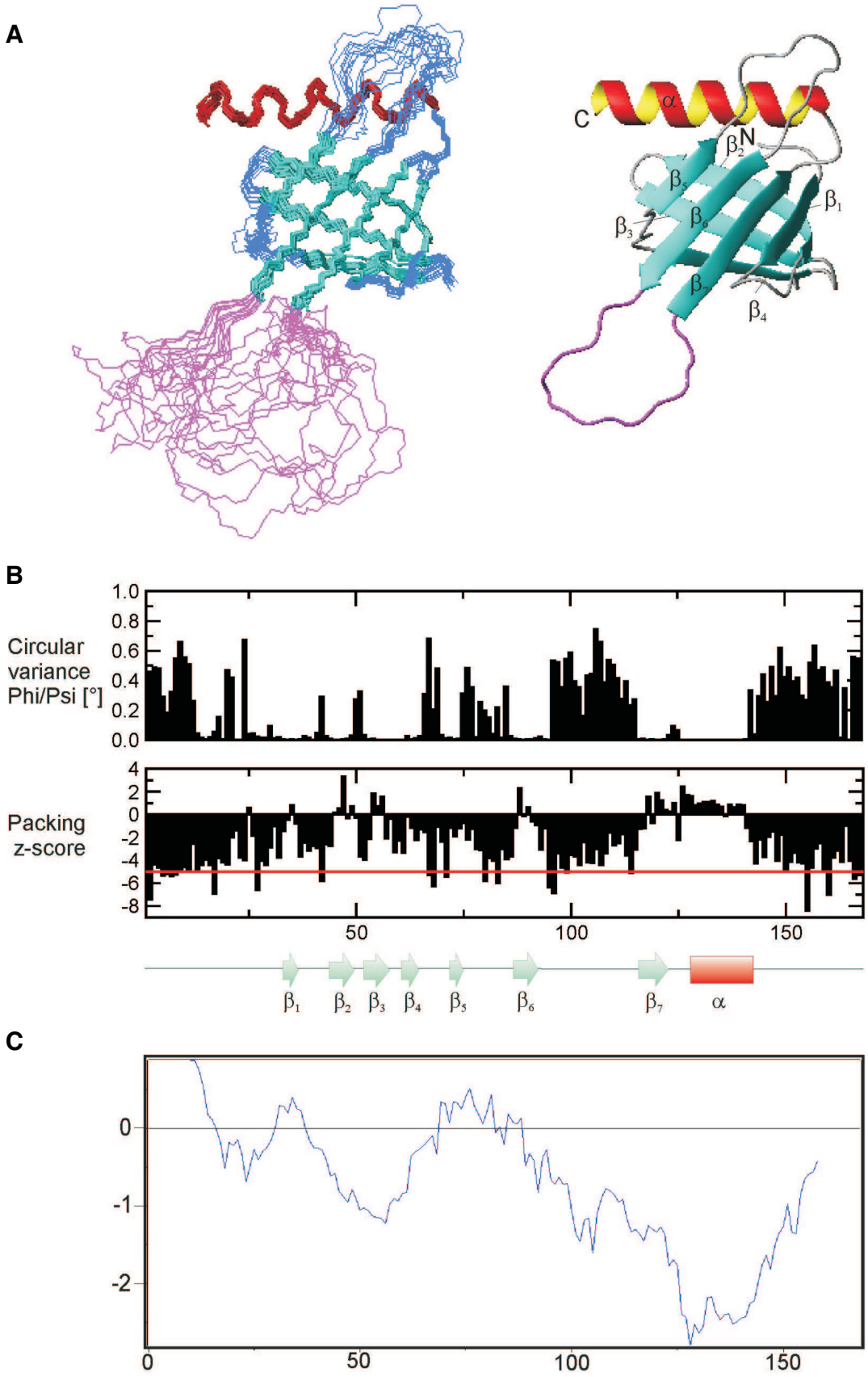


Figure 4.6: (A) Ribbon plot of the backbone structure of N-terminal recombinant $Canisfamiliaris$ ICln is drawn with β-strands in cyan and α-helices in red/yellow right to a representative ensemble of 15 models. In (B) the precision of the bundle in term of circular variances and the packing quality per residue is plotted. At bottom the secondary structure elements are given. (C) PROSA total interaction

Glu106, Glu107, and Glu114; four asparagines: Asp108, Asp110, Asp111, and Asp112). Following the seventh β-strand, a C-terminal α-helical stretch extents from residue Lys126 to residue Leu142. Its hydrophobic side facing the core (stabilizing the helix orientation on the top of the structure) is build up by residues Ala128/Leu129, Ala131/Met132, and Ala135/Met136. Its hydrophilic strip exposed to the solvent is formed by residues Lys126/Ser127, Glu130, Thr134, and Glu138.

Excluding the N-terminal His tag (starting methionine, the 6xHis-tag, and the two residue spacer Leu-Glu), about 21.4% from the remaining 159 residues of the truncated wildtype sequence (sample B) are involved in β-sheet structure, and about 10.7% are found in the C-terminal α-helical conformation. The structural statistics of the final structure ensemble containing 15 models are given in table 4.4 and table 4.5. A bundle precision over 15 respresentive models and some packing quality information thereof is given in figure 4.6.B,C.

Table 4.4: Summary of applied conformational restraints and structural statistics for the 15 fi nal models of ICln

Experimental NOE distance restraints	
total NOE distance restraints	959
intra-residue distance restraints	203
sequential backbone ($\|i-j\|$=1) distance restraints	318
inter-residue, medium-range (4$\leqq |i-j| \leqq$2) distance restraints	127
inter-residue, long-range ($|i-j| >$4) distance restraints	311
Torsions restraints	
backbone-ϕ,ψ-torsions (TALOS)	174
sidechain χ_1-torsions ($^3J_{C_\gamma N}$/$^3J_{C_\gamma C'}$)	15
Hydrogen bonds	
N-H$\cdots$O=C	48
Average RMS-deviations from	
experimental distance restraints [Å]	0.086
idealized covalent geometry	
bonds[Å]	5.337E-03 $\pm$ 1.697E-04
angles[°]	7.630E-01 $\pm$ 2.592E-02
improper[°]	5.770E-01 $\pm$ 3.815E-02
PROCHECK statistics	
Percentage residues in allowed regions of Ramachandran plot	
% in whole molecule (% in ordered structure elements[a])	
most favoured regions [A,B,L]	80.6% (95.6%)
additional allowed regions [a,b,l,p]	16.7% (4.2%)
generously allowed regions [$\sim$a, $\sim$b, $\sim$l, $\sim$p]	2.9% (0.1%)
total in allowed regions	99.5% (100%)
Omega angle standard deviation (typical value[b])	1.3 (6.0)

Experimental NOE distance restraints	
Zeta angle standard deviation (typical value[b])	0.8 (3.1)
H-bond energy standard deviation (typical value[b])	0.8 (0.8)
PROSA	
Mean force potential - whole molecule (residues in ordered structure elements[a])	
pair energy	-0.50 (-0.78)
surface energy	-0.28 (-0.34)
combined energy	-0.78 (-1.12)
z-score - whole molecule (omitting unstructured N- and C-terminus)	
pair energy	-4.03 (-3.70)
surface energy	-3.25 (-3.47)
combined energy	-4.63 (-4.84)
What if	
Overall packing score including residues	
33-142 (omitting unstructured N- and C-terminus)	-1.67
33-92, 116-142 (omitting also the highly flexible β_{67} loop)	-1.14

Table 4.5: Root mean square deviations from the mean structure calculated from the fi nal ensemble of 15 structures of ICln[a]

Selected residues	Selected atoms	RMSD [Å]
all elements of secondary structure	backbone atoms	0.60
	all heavy atoms	0.99
	all atoms	1.19
all β-strands	backbone atoms	0.35
	all heavy atoms	0.79
	all atoms	0.94

[a] statistic apply to ordered regions of ICln, with residues 33-36 (β_1),44-49 (β_2),53-56 (β_3),62-64 (β_4),72-74 (β_5),86-92 (β_6),116-122 (β_7),128-142 (C-terminal α-helix). [b] according [107].

4.1.1.5 Overall electrostatics and functional comparison of ICln with other Pleckstrin homology (PH) domain proteins

The presence of a large set of proteins at the structure database offering marginally homologous amino acid primary sequence and diverse biomolecular functions while sharing the same basic pleckstrin homology (PH) domain fold led to the notion that these domains form a *superfold* family. Although the secondary structure elements of the PHD core region are structurally conserved the connecting loops are found to be generally very variable. Furthermore, PH domains permit the insertion of other independent domains without modifying the PH fold. For instance, the PH domain of phospholipase C γ_1 contain tow Src homology 2 (SH2) domains

and one SH3 domain at loop β_3/$\beta 4$.

This is in agreement with the observation that exchanging the β_6/$\beta 7$ of 17 residue length (residue Glu89 to residue Ile107 in the original ICln wild type sequence) by a two residue (Gly-Ser) linker results in not any structural consequence for the PH domain part of ICln.

In addition, structural studies revealed that PH domains are structurally similar to other known domains such Enabled/VASP homology domain, phosphotyrosine binding domain or Ran binding domain [9].

Most of the PH domains solved by date were shown to interact with phosphorylated ligands (Inositol phosphates or phosphatidylinositol lipids, [12]). A common characteristic of these ligand binding PH domains is their electro statically polarization, whereas the Inositol phosphate ligand binding site (conservatively located at the opposite end of the β-barrel in respect to the C-terminal α-helix) is embedded by positively charged residues. Therefore it is thought that long range electrostatic effects between that positive charged surface patch and the negative charged Inositol phosphate head groups provide the orientation forces of a PH domain containing protein (as phospholipase C γ_1) towards the cell membrane. It has been shown that mutation (concerning residues outside the binding pocket) resulting in significant charge reversing at the positive potential patch led to a conspicuously decrease in ligand affinity.

The K_D values for phosphatidylinositol binding of PH domain published per date span from the nano-molar to milli-molar range. While some PH domains function as signal regulated membrane targeting modules weaker phospholipid binding PH domains (e.g. β-adrenergic receptor kinase [180]) require cooperative interactions for membrane association. PH domains are also found in cytoskeleton proteins and proteins that regulate the actin assembly [99].

Initially, ICln was also thought to bind inositol phosphates (Paulmichl and co-workers, unpublished results), wherefore titration experiments were performed (prior to structure determination) using D-*myo*-Inositol 3,4,5-trisphosphate and DL-*myo* Inositol-1,4,5,6-tetrakisphosphate suggested to be the most appropriate interaction partner for the protein. However, in both cases no chemical changes could be detected up to a molar equivalent ratio ligand to protein of 6:1.

An explanation of this finding is evidently given by comparison of the overall electrostatic of ICln to other PH domains known to bind phosphoinositols with high (phospholipase-C-δ_1) or intermediate (β-spectrin) affinity (figure 4.7). As mentioned above, the binding site opposite is enclosed by a positive patch of charges responsible for the long range interaction prior specific binding. In contrast, ICln exhibit no electro statically polarization rather a strong negative overall

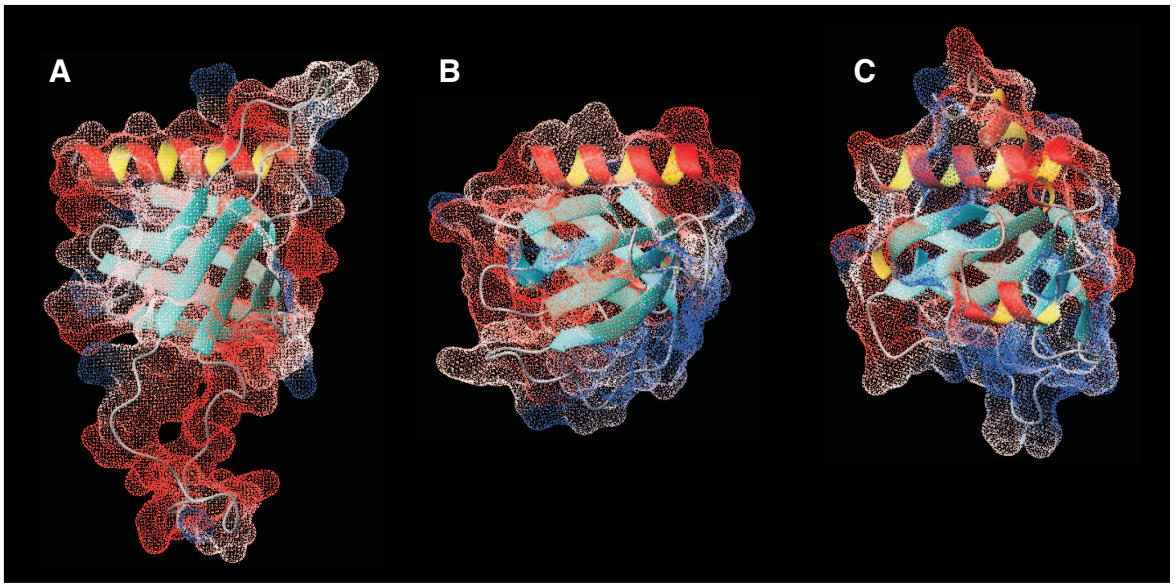


Figure 4.7: Overall electrostatic potentials comparison of (A) ICln with PH domain of (B) phospholipase-C-δ_1 (Ins(1,4,5)P_2; K_D=210 nM) and (C)β-spectrin (Ins(1,4,5)P_3; K_D=40μM). Note the accumulation of positive charges (colored in blue) enclosing the phosphoinositol ligand binding site (bottom right) in B,C.

electrostatic potential. Moreover, the β_1/β_2 loop consistently observed to form a part of the binding pocket in all phosphoinositol binding PH domains is found in close spatial neighborhood the high dynamical 17 residue loop β_6/β_7 containing the acidic domain 2 (AD2) of ICln, which can be thought to act as a repulsing 'bodyguard' against all negative charges (as phosphoinositol head groups).

In addition, the region corresponding to the phosphoinositol binding site of phospholipase C γ_1 show a more closed conformation compared to other (phospholipide binding) PH domains. In summary, these findings indicate that under hypotonic conditions ICln is translocated towards the cell membrane by other mechanism than direct phospholipid binding.

ICln shares the attribute of strong overall negative electrostatics with the more long ago published PH domain of *C.elegans* UNC89 (117 residues, [10]). Compared to all other PH domains solved per date ICln offers some structural peculiarities. On the one hand there is the more extended 6 residue loop β_4/β_5 connecting both β-sheets of the barrel compared to an usual linker length of 3 to 4 residues found in other PH domains. Beside from backbone NOE pattern analysis this is also indicated by inspection of the secondary chemical shifts. As a consequence there were only sidechain NOE contacts found between this linker region and the two β-sheets and the spatial adjacent C-terminal α helix, respectively, but not any inter-β-sheet sidechain NOEs on that site of the barrel, which led to a slightly more open conformation of the β barrel at this position. Another peculiarity of the solution structure of ICln is (as mentioned above) the

strong acidic 17 residue loop β_6/β_7 indicated to be highly dynamical from relaxation data.
In summary, the presented NMR data provide not any structural support for a channel function of ICln. Regarding the (repulsive) overall negative electrostatic and the monomere existence of ICln under all experimental conditions, which were similar to those applied for the reconstitution experiments (50mM phosphate, 150mM NaCl or 150 mM KCl, at pH 8 [46]) it remains a mystery how membrane permeability by ICln is achieved as measured by expert groups professional in bilayer and patch-clamp experiments [137], [43], [100].

4.1.1.6 Binding and stability studies of ICln

From co-immunoprecipitate experiments and affinity arrays using immobilized KVGFFKR peptides it was shown that ICln specifically interacts with the membrane-proximal regions of the platelet $\alpha_2\beta$ integrin protein and a physiologically relevant function in integrin activation was suggested [94].
The titration experiment performed at this work verified the binding capability of ICln (PH domain, ^{15}N-labeled sample B) to that intracellular part of $\alpha_2\beta$ integrin and revealed Ser71, A74, I70, and C137 (beside some other residues with lower chemical shift changes) to be part or next to the binding site (figure 4.8, table 4.6). However, that finding is in contrast to the proposed AKFEEE motif suggested by [94] to have functional significance for integrin recognition. In the tertiary structure these sequence fragment (residue Ala93 to residue Glu98, whereas in the underlying *dog* ICln wild type primary sequence the first glutamate acid is substituted by a glycine) is located in the highly dynamical loop β_6/β_7 at the opposite bottom site of ICln. With exception of Glu97 the related backbone ^{1}H^N and ^{15}N chemical shifts were assigned and show not any changes in their values upon M987-WKVGFFKRN-R997 integrin peptide addition.
Interestingly, the observed residues with notedly shifted backbone ^{1}H^N/^{15}N values form a spatial patch next to the C-terminal α-helix. As there were not any conformational changes detected for ICln upon binding interaction and the peptide was shown to exhibit a continuous α-helical conformation, the ICln/$\alpha_2\beta$ integrin peptide complex can be supposed to be a α-helix dimer as published in [193], where the solution complex structure of two synthetic peptides comprising the membrane-proximal regions of the integrin cytoplasmatic tails (M987-W*K*V*GFFK*-R995-N-R997 of α_{2b}, and *K*716-*LL*ITI*H*-*D*723-*R*KEFAKFEEERARAKW-D740 of β_3; highlighted residues = highly conserved residues, italicized residues = residues implicated in salt bridge formation) was determined.

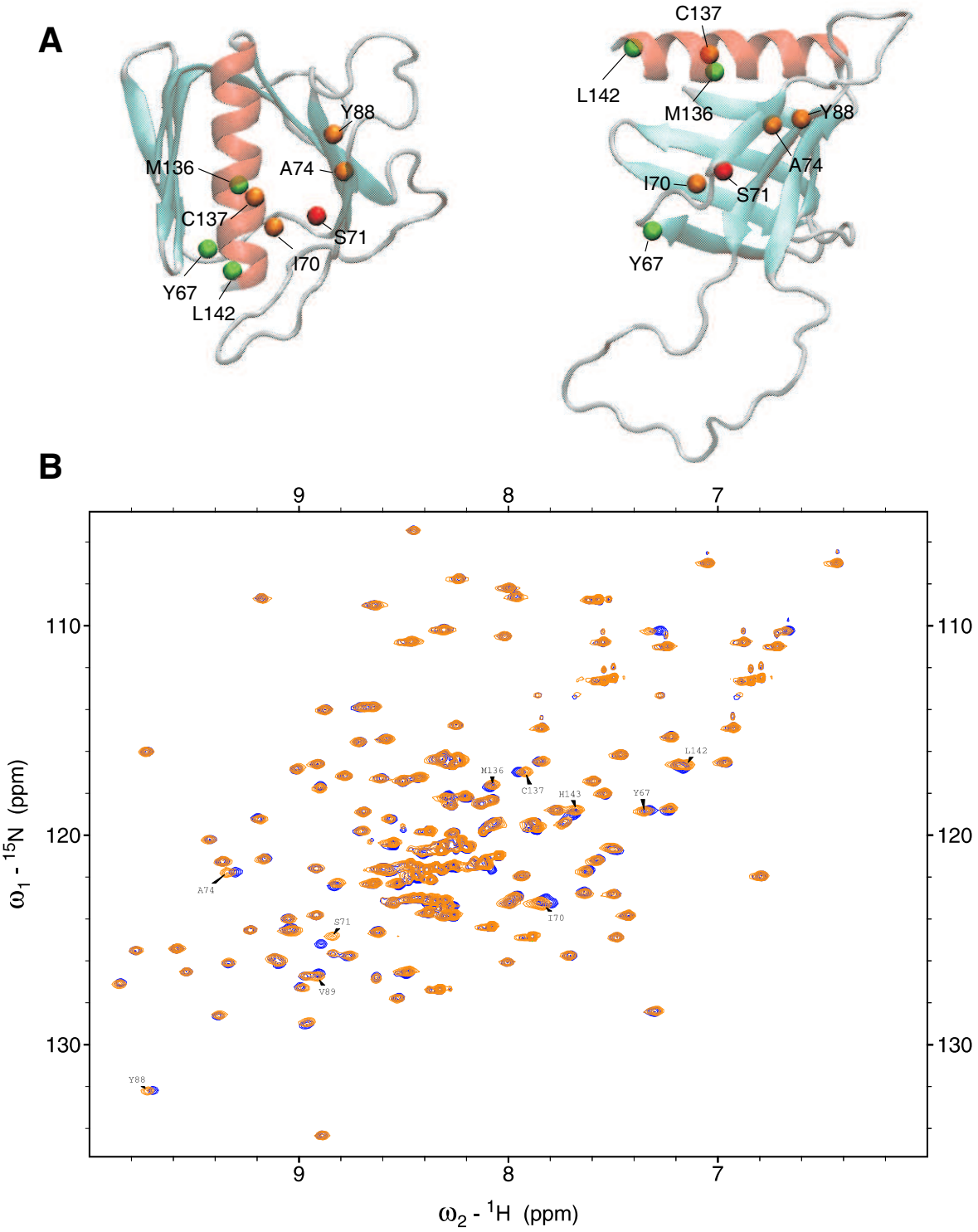


Figure 4.8: Localization of residue whose $^{1}H^{N}$ and ^{15}N chemical shifts exhibiting most signifi cant changes upon M987-WKVGFFKRN-R997 integrin peptid binding (A) at the tertiary structure of ICln (colored according extent of chemical shift alteration - ranging from red [most shifted] to orange and green) and (B) in the 2D $^{1}H^{N}$,^{15}N HSQC spectrum (indicated by black arrows; blue = apo ICln and orange = $holo$ form).

Table 4.6: Backbone chemical shift differences between *apo* ICln and the ICln/$\alpha_2\beta$ integrin peptide complex.

Residue	$(\Delta\delta^{15}\mathbf{N})^2 + (10 \cdot \Delta\delta^{1}\mathbf{H}^{N})^2$	**Location**
Tyr67	0.089	linker loop β_4/β_5
Ser71	0.436	linker loop β_4/β_5
Ala74	0.130	strand β_5
Tyr88	0.092	strand β_6
Val89	0.201	strand β_6
Met136	0.109	C-terminal α-helix
Cys137	0.108	C-terminal α-helix
Leu142	0.207	C-terminal α-helix
His143	0.205	C-terminal α-helix

As low ionic strength was needed in order to prevent precipitation of the 11 residue peptide (M987-WKVGFFKRN-R997) employed for interaction studies with ICln, MOPS instead of phosphate was used as puffer system. In the reference work [193] a complete unbuffered sample was applied to impede 'visible' precipitation of the 2mM peptide mixture. However, the softest salt condition at which ICln was stable for NMR analysis was 150 mM NaCl and 25mM MOPS at pH=7.0. Unfortunately at that state the $\alpha_2\beta$ integrin peptide successively used for titration starts to precipitate, hence preventing an exact determination of K_D values and measurement of valuable *inter* molecular NOEs to examine the exact complex structure.

As purified ICln reconstituted in lipid bilayer was shown to gave rise for a nucleoside sensitive conductance (of about 3 pS) for both cations and anions, an *in vivo* channel function was concluded (from this and other data) by [43], and a homo-dimer model was proposed as channel forming unit

Do detect possible conformational changes in hydrophobic environment ICln (sample B) was titrated with 20% solution of CHAPS (3[(3-Cholamindopropyl)dimethylammonio]-1-propane-sulfonate) starting from non-micellar concentration ($<$ 6mM $\equiv$ ciritical micelle concentration) up to 10mM. This nondenaturing zwitterionic detergent was choosen because of its small micellar molecular weight (6,150) suitable for NMR analysis. However, either assembling with micelles nor any structural alterations could be observed.

As in accordance with swelling-dependent chloride conductance found in NIH3T3 fibroblasts [44], the relative permeability of cations over anions was observed to depend on the presence of calcium. From site-directed mutagenesis an ICln calcium-binding site was suggested by [137]. Furthermore, the nucleotide sensitivity of ICln expressed in oocytes was reported to be annihilated by a single mutation at position Gly58 [43]. The reconstituted G49C mutant could

not be blocked by acyclovir, a nucleoside analogue able to impair swelling-dependent chloride channels in fibroblasts. Adding acyclovir reduces the macroscopic current $I_{Cl^-,swell}$ as well as the single-channel current induced by reconstituted ICln proteins. Basing on this the authors concluded that Gly58 is 'an amino acid crucial for binding nucleotides or nucleoside analogues to ICln'.

However, all titration experiments with Ca^{2+} and AMP (a putative binding nucleotide suggested by the Prof.Paulmichl group) revealed no specific affinity of ICln for these ligands.

The situation was different investigating Lsm4 as putative ICln interaction partner. Residues of shifted $^{1}H^{N}/^{15}N$ resonances were exclusively found to be part of the acidic domain 2 or located at the end of the C-terminal α-helix adjoining to it (see figure 4.9). One exception concerns residues Ser71 and Leu72 accommodated at the linker loop β_4/β_5 which is spatial adjacent to that region (see table 4.7).

This is in complete agreement with the findings from biochemical binding assays (2hybrid,ELISA,etc.), using several mutants which indicate the presence of the acid domain 2 as prerequisite for Lsm4 binding (Fuerst and Co-worker, unpublished).

Table 4.7: Backbone chemical shift differences observed in titration experiments with LSm4 as putative ICln ligand.

Residue	$(\Delta\delta^{15}\mathbf{N})^2 + (10 \cdot \Delta\delta^{1}\mathbf{H}^{N})^2$	**Location**
Ser71	0.035	linker loop β_4/β_5
Leu72	0.100	linker loop β_4/β_5
Ala141	0.100	C-terminal α-helix
Leu142	0.200	C-terminal α-helix
His143	0.120	C-terminal α-helix
Glu147	0.210	C-terminus, Acidic domain 2
Ser151	0.200[a]	C-terminus, Acidic domain 2

[a] overcrowded by (serine) residues of C-termini.

As mentioned before, the endurance of the ICln preparations containing the complete wild type sequence suffers due to the instability of its unfolded C-terminus. Protective osmolytes [173] as sorbitol, betaine, myo-inositol, taurine, glycerophosphocholine, proline, argenine, and glycerol are known to stabilize intra-cellular protein against a wide variety of adverse environmental conditions. For example, under hypertonic conditions cells react against osmotic shrinkage by regulatory volumes increase (RVI, the counterpart of RVD described in chapter 1) in which transporters are activated that mediate rapid uptake of inorganic ions, followed by an osmotic influx of water that restores cell volume back towards normal within minutes. As a side effect

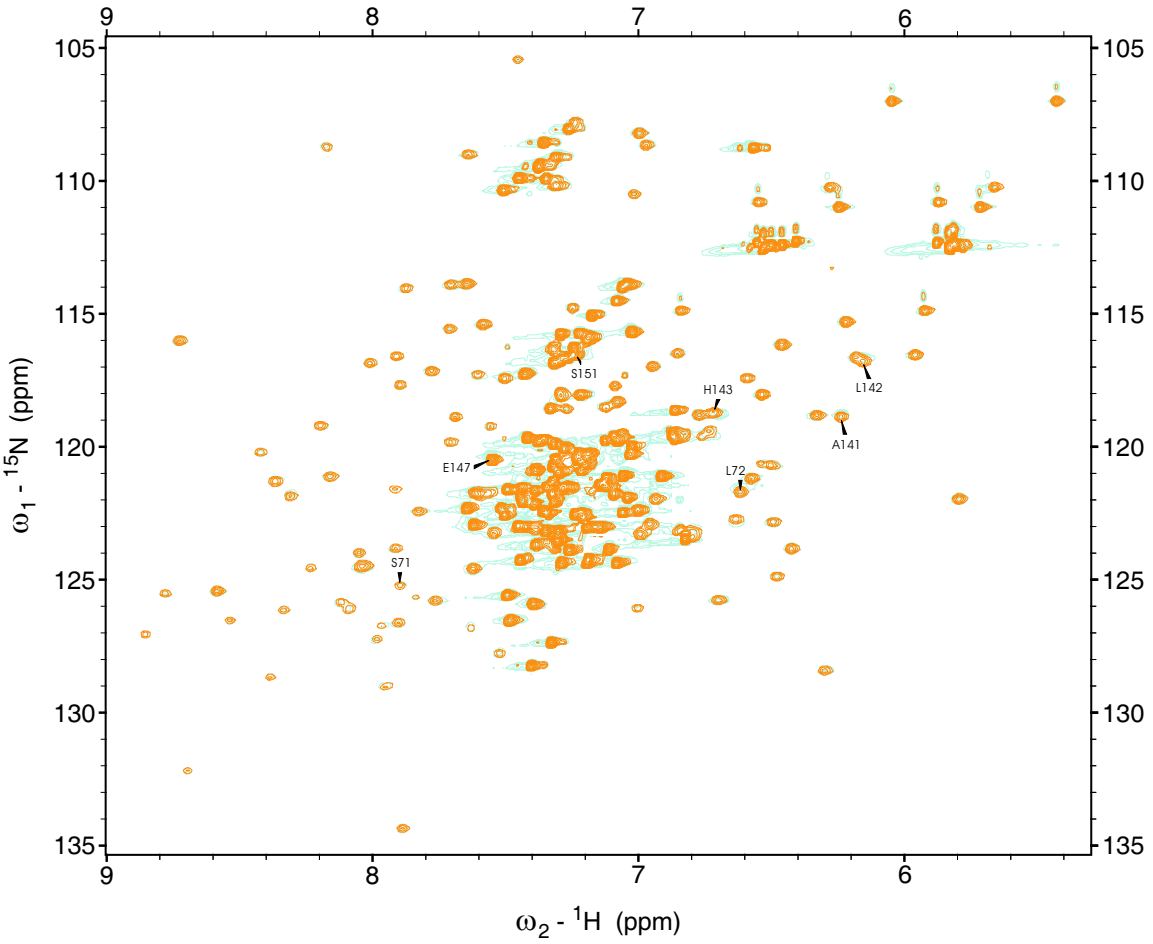


Figure 4.9: 2D $^{1}H^{N}$,^{15}N HSQC spectra overlay of wild type ICln before (orange, sample A) and upon Lsm4 addition (light blue). Peak positions of residues with chemical shift changes are labeled.

intracellular ionic strength I perturbing intracellular macromolecules strongly increases. However, over a period of hours, protective osmolytes accumulate in cells and gradually reducing I back towards that in the normotonic state while maintaining cell volume. Because of some interest (Fuerst and co-workers) the extent of stability increasement of full length ICln (complete 235 residues) in the presence of the osmolyte L-argenine was evaluated. While an original ICln preparation was found to be stable for about five days, the availability of mM L-argenine increases the protein stability to almost three weeks. Until this time period not any chemical shift changes or occurrence of resonances due to degeneration were observed.

4.1.1.7 Phosphorylation state of ICln

As ICln offers a number of consensus sites for phosphorylation by protein kinase C (beside protein kinase A and G; casein kinase I and II), the phosphorylation state of an uniformly ^{15}N-labeled sample containing the PH domain of wt ICln (sample B1) after *in vitro* PKC treatment

was analyzed. However, not any chemical shift changes could be detected using the provided sample. Subsequent analysis [45] revealed that ICln are substrate for protein kinase A (PKA), PKC, and protein kinase γ_1 (PKγ_1), whereas Ser54 was reported to be the sole target for PKA and the main target for PKC as well as PKγ_1. Furthermore, mass spectra analysis indicated N-terminal Ser11 (Ser2 in numbering without starting methionine and His tag) and Ser102 (corresponding Ser93, accommodated at loop β_6/β_7) as additional PKγ_1 phosphorylation sites. Threonine or tyrosine phosphorylation was not detected.

4.1.2 Cyclophilin D

4.1.2.1 Resonance assignment

Cyclophilin D gave rise for excellent NMR spectra with a surpassingly ^{1}H-, ^{15}N-, and ^{13}C chemical shift dispersion leading to an essentially complete backbone sequential resonance assignment as demonstrated by the ^{1}H-^{15}N HSQC spectrum (figure 4.10). Together with the obtained remarkable high signal/noise ratio the quality of the triple resonance spectra facilitates an automated performance of the backbone assignment process, which was initially performed and provided by Bernd Hoffmann [65] for structure determination using a simulated annealing Monte Carlo algorithm analog to the MONTE program [63]. In summary, 154 of the 158 non-proline backbone amide resonances were assigned. The backbone H$^{\alpha}$ protons were extracted manually from the 3D ^{13}C edited NOESY experiment.

With exception of the N-terminal Met1, His70, Met100, Asn102, and Ser 149, for which no assignment was possible due to signal absence in the NMR spectra (presumably because of their rapidly exchanging amide protons), an almost complete backbone assignment serves as starting point for subsequent sidechain chemical shift and NOESY assignment. In contrast to the case of ICln the 3D H(CCCO)NH and 3D (H)CC(CO)NH spectra were very useful for that purpose.

In total, 93.5% of the aliphatic- and 85.1% of the aromatic sidechain ^{1}H,^{13}C chemical shifts could be assigned. Assignment of aromatic protons (involved in valuable protein core NOEs) was accomplished for all 11 phenylalanines and for 2 (out of 3) tyrosines. The achieved sidechain assignment level for the methyl bearing residues was 87.4%, whereas a complete sidechain assignment for all 20 leucines, 14 isoleucines, 8 alanines, and 4 threonines were obtained. For residue type valine, 13 (out of 15) methyl functions could be assigned. Due to some ^{1}H- and ^{13}C resonance overlap in the 3D ^{13}C edited NOESY spectrum methyl function

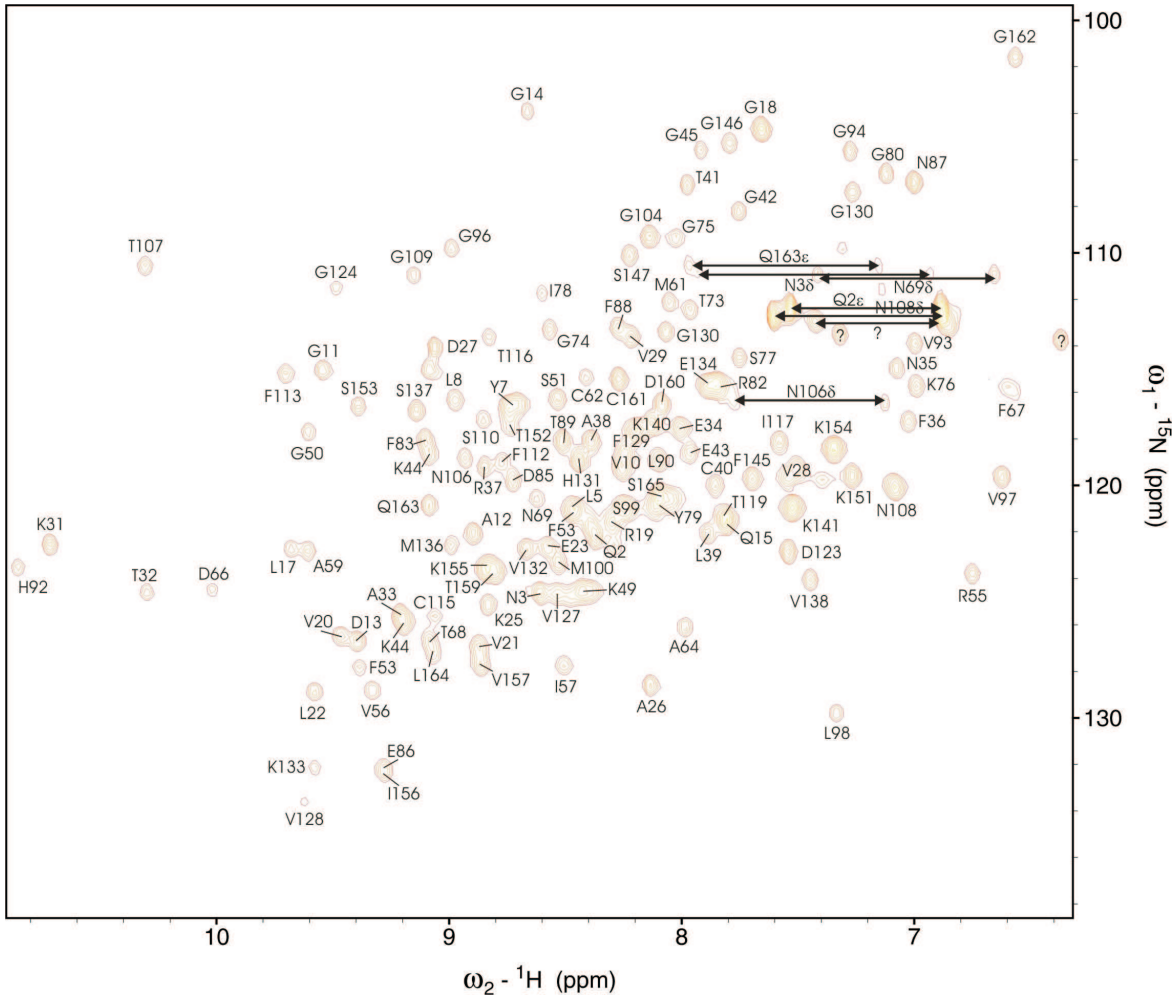


Figure 4.10: 800 MHz 2D $^{1}H^{N}$,^{15}N HSQC of cyclophilin D.

assignment of neither Met61 nor Met136 was possible.

In addition, ^{1}H,^{15}N chemical shift assignment for 8 (out of 14 observable) sidechain amide functions was possible (for 3 asparagine δ-NH_2 and for 5 ϵ-NH_2 glutamine groups, in total 57.1%). Furthermore, the $N^{\epsilon 1}/H^{N\epsilon 1}$ resonances of the only tryptophane residue (Trp121) could be unambiguously assigned.

The backbone and sidechain chemical shift assignments of wild type cyclophilin D have been deposited in the BioMagResBank under BMRB accession code number 7810 (see also table 4.8). Because of the high number of aromatic residues in cyclophilin D numerous ^{1}H,^{13}C sidechain signals with large ring current shifts were found.

Table 4.8: Backbone and sidechain chemical shift assignment of cyclophilin D

Residue	δ(**H**^{N}) [ppm]	δ(^{15}**N**) [ppm]	δ(**H**$_\alpha$/**C**$_\alpha$) [ppm]	**others** [ppm]
MET1	-	-	4.43/55.05	33.05(C_β); 31.05(C_γ);
GLN2	8.39	122.15	4.31/55.06	2.04,2.04/29.91(H_β/C_β); 2.33,2.33/33.70(H_γ/C_γ); 7.62,6.89/112.78(H_δ/N_δ)
ASN3	8.62	124.59	5.09/51.25	3.54,2.64/37.88(H_β/C_β); 7.94,6.95/110.95(H_δ/N_δ)
PRO4	-	-	4.67	2.30,1.72/32.98(H_β/C_β); 2.15,2.11/26.18(H_γ/C_γ); 4.11,4.01/(H_δ/C_δ)
LEU5	8.48	121.03	5.52/53.04	1.89,1.24/44.78(H_β/C_β); 1.53/26.90(H_γ/C_γ); 0.88/25.44($H_{\delta1}$/$C_{\delta1}$); 0.86/24.28($H_{\delta2}$/$C_{\delta2}$)
VAL6	9.11	115.02	5.37/58.09	2.65/35.09(H_β/C_β); 0.92/21.85($H_{\gamma1}$/$C_{\gamma1}$); 1.12/20.20($H_{\gamma2}$/$C_{\gamma2}$)
TYR7	8.72	116.45	6.15/54.85	2.78,2.54/42.63(H_β/C_β); 6.57(H_δ); 6.40(H_ϵ)
LEU8	8.99	116.24	4.70/53.73	1.93,1.37/45.06(H_β/C_β); 0.95/26.22($H_{\delta1}$/$C_{\delta1}$); 0.90/23.21($H_{\delta2}$/$C_{\delta2}$)
ASP9	9.21	125.91	5.63/54.48	2.58,2.50/41.66(H_β/C_β)
VAL10	8.27	119.30	5.51/54.87	1.55/34.28(H_β/C_β); 0.88/21.37($H_{\gamma1}$/$C_{\gamma1}$); 0.88/21.37($H_{\gamma2}$/$C_{\gamma2}$)
GLY11	9.56	115.02	4.99,3.61/44.48	-
ALA12	8.92	122.04	5.17/49.98	0.75/21.61 (H_β/C_β)
ASP13	9.40	126.73	4.36/55.64	3.10,2.45/39.91(H_β/C_β)
GLY14	8.67	103.96	4.17,3.61/45.45	-
GLN15	7.83	121.43	4.85/51.86	2.18,2.03/29.52(H_β/C_β); 2.37,2.29/33.31(H_γ/C_γ); 7.55,6.90/112.17(H_δ/N_δ)
PRO16	-	-	4.45	2.28,1.91/32.53(H_β/C_β); 26.71(C_γ);
LEU17 LEU17	9.70	122.76	4.59/54.87	1.67,1.41/43.41(H_β/C_β); 1.68/26.80(H_γ/C_γ); 0.93/22.34($H_{\delta1}$/$C_{\delta1}$); 0.74/25.74($H_{\delta2}$/$C_{\delta2}$)
GLY18	7.67	104.67	4.62,3.73/44.09	-
ARG19 LEU17	8.32	121.53	5.51/54.87	1.96,1.60/34.28(H_β/C_β); 1.53,1.34/27.39(H_γ/C_γ); 2.78,2.64/(H_δ/C_δ)
VAL20	9.48	126.42	4.60/60.98	2.06/34.38(H_β/C_β);

Residue	δ(**H**^{N}) [ppm]	δ(^{15}**N**) [ppm]	δ(**H**$_\alpha$/**C**$_\alpha$) [ppm]	**others** [ppm]
				0.95/20.98($H_{\gamma1}$/$C_{\gamma1}$); 0.95/20.98($H_{\gamma2}$/$C_{\gamma2}$)
VAL21	8.89	126.93	4.94/61.66	1.92/33.50(H_β/C_β); -/-($H_{\gamma1}$/$C_{\gamma1}$); 0.98/21.46($H_{\gamma2}$/$C_{\gamma2}$)
LEU22	9.58	128.86	4.94/53.36	1.62,0.94/43.12(H_β/C_β); 0.86/20.98($H_{\delta1}$/$C_{\delta1}$); 0.86/20.98($H_{\delta2}$/$C_{\delta2}$)
GLU23	8.58	122.55	4.95/54.77	1.68,1.68/31.85(H_β/C_β); 2.08,1.75/36.90(H_γ/C_γ)
LEU24	8.40	124.49	4.88/52.07	1.97,1.31/42.88(H_β/C_β); 32.47(C_γ); 1.22/27.00($H_{\delta1}$/$C_{\delta1}$); 0.52/22.73($H_{\delta2}$/$C_{\delta2}$)
LYS25	8.85	125.20	4.52/53.60	1.65,1.58/27.39(H_β/C_β); 1.33,1.23/23.41(H_γ/C_γ); 2.89,2.80/41.56(H_ϵ/C_ϵ)
ALA26	8.15	128.56	3.70/54.19	1.45/18.74 (H_β/C_β)
ASP27	9.08	114.00	4.19/55.16	2.77,2.68/38.46(H_β/C_β)
VAL28	7.55	119.80	4.18/63.90	1.91/35.25(H_β/C_β); 1.17/20.69($H_{\gamma1}$/$C_{\gamma1}$); 1.17/20.69($H_{\gamma2}$/$C_{\gamma2}$)
VAL29	8.24	113.60	4.58/57.29	2.53/31.37(H_β/C_β);
PRO30	-	-	4.36	2.61,1.96/31.37(H_β/C_β); 26.62(C_γ); 3.60,3.42/(H_δ/C_δ)
LYS31	10.73	122.55	4.03/60.21	1.67,1.42/31.66(H_β/C_β); 1.37,1.18/25.93(H_γ/C_γ); 2.07,2.07/(H_δ/C_δ); 41.13(C_ϵ)
THR32	10.31	124.59	-/60.21	3.95/67.59(H_β/C_β); 0.88/21.17(H_γ/C_γ)
ALA33	9.22	125.50	4.10/55.84	1.44/18.65 (H_β/C_β)
GLU34	8.02	117.57	4.49/57.97	2.25,1.72/28.16(H_β/C_β); 2.46,2.01/33.70(H_γ/C_γ)
ASN35	7.09	114.92	4.00/56.61	2.84,2.31/39.62(H_β/C_β);
PHE36	7.04	117.16	4.03/61.96	3.18,3.12/40.50(H_β/C_β); 7.08(H_δ); 7.18(H_ϵ); 7.18(H_ζ)
ARG37 ARG37	8.86	119.19	3.61/60.60	1.93,1.82/30.20(H_β/C_β); 1.48,1.00/24.47(H_γ/C_γ); 3.55,2.78/(H_δ/C_δ)
ALA38	8.40	118.07	4.07/54.19	1.20/18.26 (H_β/C_β)
LEU39	7.91	122.04	3.80/57.29	1.45,-0.29/40.88(H_β/C_β); 0.88/26.71($H_{\delta1}$/$C_{\delta1}$); 0.53/24.28($H_{\delta2}$/$C_{\delta2}$)
CYS40	7.86	120.01	4.39/63.70	3.29,2.15/26.90(H_β/C_β)

Residue	δ(**H**^{N}) [ppm]	δ(**^{15}N**) [ppm]	δ(**H$_\alpha$/C$_\alpha$**) [ppm]	**others** [ppm]
THR41	7.99	107.08	4.74/63.17	4.44/69.14(H$_\beta$/C$_\beta$); 1.30/22.63(H$_\gamma$/C$_\gamma$)
GLY42	7.78	108.30	3.83,3.51/45.54	-
GLU43	7.98	118.58	4.12/58.56	2.20,2.02/30.11(H$_\beta$/C$_\beta$); 2.08,1.79/35.35(H$_\gamma$/C$_\gamma$)
LYS44	9.10	118.68	4.31/54.28	0.93,1.64/30.40(H$_\beta$/C$_\beta$); 1.04,1.03/24.86(H$_\gamma$/C$_\gamma$); 1.44,1.44/(H$_\delta$/C$_\delta$); 2.89,2.78/41.64(H$_\epsilon$/C$_\epsilon$)
GLY45	7.93	105.59	4.32,3.55/44.77	-
PHE46	6.39	113.70	4.64/54.19	3.12,2.68/39.72(H$_\beta$/C$_\beta$); 6.62(H$_\delta$); 7.02(H$_\epsilon$); 7.02(H$_\zeta$)
GLY47	7.68	104.58	4.26,2.76/45.84	-
TYR48	6.86	113.19	4.36/57.88	3.11,3.01/38.84(H$_\beta$/C$_\beta$); 6.87(H$_\delta$); 6.87(H$_\epsilon$)
LYS49	8.44	124.59	3.71/61.08	2.01,1.65/31.47(H$_\beta$/C$_\beta$); 1.33,1.30/24.67(H$_\gamma$/C$_\gamma$); 1.95,1.81/(H$_\delta$/C$_\delta$); 3.28,3.09/42.15(H$_\epsilon$/C$_\epsilon$)
GLY50	9.62	117.67	4.45,3.61/45.54	-
SER51	8.55	116.24	4.94/58.46	64.66(C$_\beta$)
THR52	9.52	107.29	5.86/61.37	4.32/73.41(H$_\beta$/C$_\beta$); 1.21/22.43(H$_\gamma$/C$_\gamma$)
PHE53	8.48	121.03	4.91/56.91	2.89,2.51/38.84(H$_\beta$/C$_\beta$); 6.89(H$_\delta$); 6.89(H$_\epsilon$); 6.89(H$_\zeta$)
HIS54	7.39	119.80	4.74/56.98	3.24,2.76/31.76(H$_\beta$/C$_\beta$)
ARG55	6.77	123.67	4.93/54.77	1.81,1.75/32.05(H$_\beta$/C$_\beta$); 1.56,1.07/28.16(H$_\gamma$/C$_\gamma$); 3.14,3.01/(H$_\delta$/C$_\delta$)
VAL56	9.35	128.66	4.47/61.96	1.80/35.03(H$_\beta$/C$_\beta$); 1.05/22.82(H$_{\gamma1}$/C$_{\gamma1}$); 0.74/21.37(H$_{\gamma2}$/C$_{\gamma2}$)
ILE57	8.51	127.74	5.19/57.49	1.88/40.88(H$_\beta$/C$_\beta$); 1.76,1.65/28.94(H$_{\gamma1}$/C$_{\gamma1}$); 1.16/17.39(H$_{\gamma2}$/C$_{\gamma2}$); 0.93/13.50(H$_\delta$/C$_\delta$)
PRO58	-	-	4.25	2.23,1.75/32.04(H$_\beta$/C$_\beta$); 1.91,1.83/27.10(H$_\gamma$/C$_\gamma$); 3.89,3.35/(H$_\delta$/C$_\delta$)
ALA59	9.62	122.76	3.81/52.92	1.40/16.41 (H$_\beta$/C$_\beta$)
PHE60	7.44	112.78	4.99/57.49	3.32,2.94/38.07(H$_\beta$/C$_\beta$); 7.19(H$_\delta$); 7.19(H$_\epsilon$); 7.19(H$_\zeta$)
MET61	8.08	112.07	5.17/54.67	2.20,1.44/35.35(H$_\beta$/C$_\beta$); 1.02,1.02/28.36(H$_\gamma$/C$_\gamma$);
CYS62	8.43	115.33	4.84/56.56	3.14,2.72/30.11(H$_\beta$/C$_\beta$)
GLN63	9.40	127.84	5.21/54.87	2.08,1.88/30.79(H$_\beta$/C$_\beta$);

Residue	δ(**H**^{N}) [ppm]	δ(**^{15}N**) [ppm]	δ(**H$_\alpha$/C$_\alpha$**) [ppm]	**others** [ppm]
				1.31,1.04/32.84(H$_\gamma$/C$_\gamma$);
ALA64	7.99	126.11	4.78/50.05	-0.21/22.53 (H$_\beta$/C$_\beta$)
GLY65	-	-	-	-
ASP66	10.03	124.49	4.34/51.47	2.72,1.67/38.75(H$_\beta$/C$_\beta$)
PHE67	6.62	115.73	4.64/56.13	2.72,2.72/38.87(H$_\beta$/C$_\beta$); 7.20(H$_\delta$); 7.20(H$_\epsilon$); 7.20(H$_\zeta$)
THR68	7.33	109.83	4.53/61.86	4.02/68.75(H$_\beta$/C$_\beta$); 0.73/22.05(H$_\gamma$/C$_\gamma$)
ASN69	8.63	120.52	4.76/52.73	2.85,2.60/39.33(H$_\beta$/C$_\beta$); 7.43,6.66/110.95(H$_\delta$/N$_\delta$)
HIS70	-	-	-	-
ASN71	7.59	112.58	4.70/52.54	2.70,2.70/38.79(H$_\beta$/C$_\beta$);
GLY72	9.67	110.64	4.58,3.35/45.16	-
THR73	7.98	112.48	4.44/62.54	4.25/70.99(H$_\beta$/C$_\beta$); 1.07/20.98(H$_\gamma$/C$_\gamma$)
GLY74	8.59	113.29	4.48,3.55/45.64	-
GLY75	8.04	109.32	4.55,2.57/43.11	-
LYS76	7.01	115.73	4.60/56.13	1.86,1.54/34.67(H$_\beta$/C$_\beta$); 22.68(C$_\gamma$); 1.15,1.15/(H$_\delta$/C$_\delta$); 40.44(C$_\epsilon$)
SER77	7.78	114.41	5.23/56.81	4.23,4.23/69.63(H$_\beta$/C$_\beta$)
ILE78	8.61	111.76	4.20/63.70	1.70/37.29(H$_\beta$/C$_\beta$); 0.62,-0.40/23.02(H$_{\gamma 1}$/C$_{\gamma 1}$); 0.70/17.48(H$_{\gamma 2}$/C$_{\gamma 2}$); 0.45/14.08(H$_\delta$/C$_\delta$)
TYR79	8.07	120.62	4.73/56.04	3.42,2.35/38.84(H$_\beta$/C$_\beta$); 6.42(H$_\delta$); 6.13(H$_\epsilon$)
GLY80	7.14	106.57	4.67,3.83/43.99	-
ARG82	7.86	115.63	5.56/53.70	1.60,1.60/34.28(H$_\beta$/C$_\beta$); 1.51,1.44/26.90(H$_\gamma$/C$_\gamma$); 3.15,2.90/(H$_\delta$/C$_\delta$)
PHE83	9.12	117.87	5.21/54.96	3.11,2.97/39.43(H$_\beta$/C$_\beta$); 7.28(H$_\delta$); 7.36(H$_\epsilon$); 7.36(H$_\zeta$)
PRO84	-	-	4.00/-	2.24,1.70/32.82(H$_\beta$/C$_\beta$); 1.98,1.89/27.39(H$_\gamma$/C$_\gamma$); 3.83,3.72/(H$_\delta$/C$_\delta$)
ASP85	8.74	119.70	4.14/54.77	2.36,1.88/40.88(H$_\beta$/C$_\beta$)
GLU86	9.30	131.92	3.79/60.40	2.21,1.79/31.65(H$_\beta$/C$_\beta$); 2.39,2.30/34.96(H$_\gamma$/C$_\gamma$)
ASN87	7.02	106.98	4.12/52.83	3.37,2.73/39.52(H$_\beta$/C$_\beta$); 7.84,6.58/116.14(H$_\delta$/N$_\delta$)
PHE88	8.29	113.29	5.97/56.03	3.56,2.55/38.36(H$_\beta$/C$_\beta$); 7.26(H$_\delta$); 7.26(H$_\epsilon$); 7.26(H$_\zeta$)
THR89	8.52	118.07	3.73/66.81	3.74/69.04(H$_\beta$/C$_\beta$); 1.17/21.95(H$_\gamma$/C$_\gamma$)

Residue	δ(**H**^{N}) [ppm]	δ(^{15}**N**) [ppm]	δ(**H**$_\alpha$/**C**$_\alpha$) [ppm]	**others** [ppm]
LEU90	8.12	118.89	4.46/54.48	1.77,1.46/40.88(H_β/C_β); 1.61/28.26(H_γ/C_γ); 21.33($H_{\delta1}$/$C_{\delta1}$); 0.80/25.54($H_{\delta2}$/$C_{\delta2}$)
LYS91	8.26	118.58	4.66/54.09	2.21,1.60/34.28(H_β/C_β); 1.46,1.39/25.64(H_γ/C_γ); 1.62,1.56/(H_δ/C_δ); 3.04,3.04/42.63(H_ϵ/C_ϵ)
HIS92	10.87	123.47	4.38/56.91	3.05,2.85/25.83(H_β/C_β)
VAL93	7.02	113.90	3.99/63.41	2.34/32.73(H_β/C_β); 0.97/22.24($H_{\gamma1}$/$C_{\gamma1}$); 0.83/17.77($H_{\gamma2}$/$C_{\gamma2}$)
GLY94	7.30	105.61	4.38,3.63/45.74	-
PRO95	-	-	3.75/-	1.96,1.84/31.85(H_β/C_β); -,-/26.81(H_γ/C_γ); 3.60,3.42/(H_δ/C_δ)
GLY96	9.00	109.73	4.46,3.24/45.06	-
VAL97	6.64	119.60	3.63/64.48	2.02/33.50(H_β/C_β); 0.95/24.28($H_{\gamma1}$/$C_{\gamma1}$); 1.36/24.67($H_{\gamma2}$/$C_{\gamma2}$)
LEU98	7.35	129.78	4.80/50.15	0.51,0.26/44.09(H_β/C_β); -0.03/25.74(H_γ/C_γ); 0.30/24.18($H_{\delta1}$/$C_{\delta1}$); -0.50/27.25($H_{\delta2}$/$C_{\delta2}$)
SER99	8.25	121.03	5.21/55.45	2.77,2.77/64.97(H_β/C_β)
MET100	8.54	123.37	5.34/53.60	2.76,2.41/31.95(H_β/C_β); 2.58,2.34/31.17(H_γ/C_γ);
ASN102	-	-	4.77	3.29,2.92/40.40(H_β/C_β);
ALA103	8.83	123.26	4.79/50.29	1.31/19.23 (H_β/C_β)
GLY104	8.15	109.22	4.61,3.70/43.51	-
PRO105	-	-	4.36	2.32,1.80/31.95(H_β/C_β); 27.02(C_γ); 3.61,3.52/(H_δ/C_δ)
ASN106	8.95	118.79	3.99/54.48	3.08,2.63/37.00(H_β/C_β); 7.79,7.14/116.55(H_δ/N_δ)
THR107	10.33	110.54	4.39/60.40	4.41/68.95(H_β/C_β); 0.88/21.17(H_γ/C_γ)
ASN108	7.10	120.11	4.22/55.64	1.51,0.99/39.82(H_β/C_β); 7.57,5.95/112.58(H_δ/N_δ)
GLY109	9.16	110.85	4.60,3.64/45.93	-
SER110	8.86	117.16	4.77/57.53	65.75(C_β)
GLN111	8.48	124.79	5.28/57.78	2.23,2.11/32.05(H_β/C_β); 2.45,1.86/35.64(H_γ/C_γ);
PHE112	8.79	118.89	5.90/55.35	3.33,3.25/43.12(H_β/C_β); 6.79(H_δ); 6.99(H_ϵ); 6.99(H_ζ)
PHE113	9.71	115.22	5.71/54.87	2.72,2.72/43.89(H_β/C_β);

Residue	δ(**H**^{N}) [ppm]	δ(**^{15}N**) [ppm]	δ(**H$_\alpha$/C$_\alpha$**) [ppm]	**others** [ppm]
				6.74(H$_\delta$); 6.74(H$_\epsilon$); 6.74(H$_\zeta$)
ILE114	8.76	117.36	5.01/59.33	1.59/40.88(H$_\beta$/C$_\beta$); 1.87,0.85/28.16(H$_{\gamma1}$/C$_{\gamma1}$); 0.73/18.07(H$_{\gamma2}$/C$_{\gamma2}$); 0.95/13.99(H$_\delta$/C$_\delta$)
CYS115	9.06	125.50	4.76/60.34	1.85,0.94/28.94(H$_\beta$/C$_\beta$)
THR116	8.84	113.60	4.39/61.08	4.38/67.59(H$_\beta$/C$_\beta$); 0.81/23.31(H$_\gamma$/C$_\gamma$)
ILE117	7.59	118.07	4.35/59.82	1.81/44.09(H$_\beta$/C$_\beta$); 1.35,1.26/27.10(H$_{\gamma1}$/C$_{\gamma1}$); 1.12/15.05(H$_{\gamma2}$/C$_{\gamma2}$); 0.92/17.68(H$_\delta$/C$_\delta$)
LYS118	8.27	120.92	3.76/57.39	1.98,1.65/32.92(H$_\beta$/C$_\beta$); 1.60,1.60/24.57(H$_\gamma$/C$_\gamma$); 1.76,1.75/(H$_\delta$/C$_\delta$); 3.00,3.00/41.18(H$_\epsilon$/C$_\epsilon$)
THR119	7.83	121.33	3.56/57.29	2.91/67.88(H$_\beta$/C$_\beta$); 1.03/20.59(H$_\gamma$/C$_\gamma$)
ASP120	8.41	121.74	3.54/57.10	39.60(C$_\beta$)
TRP121	7.37	118.38	5.48/59.92	3.40,3.33/26.80(H$_\beta$/C$_\beta$); 9.98(H$_{\epsilon1}$/N$_{\epsilon1}$)
LEU122	7.11	119.80	4.29/54.28	1.28,0.24/38.94(H$_\beta$/C$_\beta$); 1.67/29.33(H$_\gamma$/C$_\gamma$); 0.43/25.06(H$_{\delta1}$/C$_{\delta1}$); 0.84/21.56(H$_{\delta2}$/C$_{\delta2}$)
ASP123	7.56	122.76	5.21/55.93	2.89,2.74/39.43(H$_\beta$/C$_\beta$)
GLY124	9.50	111.46	3.90,2.84/44.96	-
LYS125	7.88	115.63	4.14/56.42	1.82,1.75/35.35(H$_\beta$/C$_\beta$); 1.26,1.21/25.25(H$_\gamma$/C$_\gamma$); 1.67,1.67/(H$_\delta$/C$_\delta$); 3.01,3.01/42.44(H$_\epsilon$/C$_\epsilon$)
HIS126	7.52	119.19	4.76/54.88	3.96,3.22/31.66(H$_\beta$/C$_\beta$)
VAL127	8.55	124.59	4.21/63.80	1.98/33.50(H$_\beta$/C$_\beta$); 1.12/20.78(H$_{\gamma1}$/C$_{\gamma1}$); 1.12/20.78(H$_{\gamma2}$/C$_{\gamma2}$)
VAL128	9.64	133.54	4.08/62.83	1.65/30.88(H$_\beta$/C$_\beta$); 0.78/22.73(H$_{\gamma1}$/C$_{\gamma1}$); -0.02/20.01(H$_{\gamma2}$/C$_{\gamma2}$)
PHE129	8.21	117.69	4.78/50.98	3.09,2.48/41.95(H$_\beta$/C$_\beta$); 6.59(H$_\delta$); 6.84(H$_\epsilon$); 6.84(H$_\zeta$)
GLY130	7.28	107.39	4.08,4.00/46.81	-
HIS131	8.46	118.89	5.10/55.64	3.09,2.80/34.67(H$_\beta$/C$_\beta$)
VAL132	8.68	122.76	3.91/64.19	1.96/32.44(H$_\beta$/C$_\beta$); 0.92/22.73(H$_{\gamma1}$/C$_{\gamma1}$); 0.73/22.24(H$_{\gamma2}$/C$_{\gamma2}$)
LYS133	9.59	131.95	4.48/56.61	1.69,1.65/35.06(H$_\beta$/C$_\beta$);

Residue	δ(**H**^{N}) [ppm]	δ(**^{15}N**) [ppm]	δ(**H$_\alpha$/C$_\alpha$**) [ppm]	**others** [ppm]
				1.52,1.48/25.06(H$_\gamma$/C$_\gamma$); 1.76,1.65/(H$_\delta$/C$_\delta$); 3.03,3.03/41.95(H$_\epsilon$/C$_\epsilon$)
GLU134	7.90	115.63	4.75/55.08	2.10,1.77/33.60(H$_\beta$/C$_\beta$); 2.27,2.17/35.93(H$_\gamma$/C$_\gamma$)
GLY135	8.85	109.12	4.92,4.15/46.13	-
MET136	9.01	122.55	4.50/56.71	2.10,2.02/29.81(H$_\beta$/C$_\beta$); 1.96,1.96/32.05(H$_\gamma$/C$_\gamma$);
ASP137	9.15	116.85	4.21/56.71	2.73,2.57/39.04(H$_\beta$/C$_\beta$)
VAL138	7.47	123.98	3.38/66.03	2.38/30.88(H$_\beta$/C$_\beta$); 1.01/21.46(H$_{\gamma1}$/C$_{\gamma1}$); 0.55/20.59(H$_{\gamma2}$/C$_{\gamma2}$)
VAL139	7.54	120.92	3.30/67.00	2.53/30.88(H$_\beta$/C$_\beta$); 1.03/23.79(H$_{\gamma1}$/C$_{\gamma1}$); 0.73/21.95(H$_{\gamma2}$/C$_{\gamma2}$)
LYS140	8.14	117.36	4.15/57.97	2.54,2.38/30.88(H$_\beta$/C$_\beta$); 1.65,1.55/24.18(H$_\gamma$/C$_\gamma$); 1.67,1.59/(H$_\delta$/C$_\delta$); 2.94,2.85/41.37(H$_\epsilon$/C$_\epsilon$)
LYS141	7.54	120.98	3.95/59.43	2.12,1.88/32.73(H$_\beta$/C$_\beta$); 1.70,1.49/25.54(H$_\gamma$/C$_\gamma$); 1.76,1.63/(H$_\delta$/C$_\delta$); 2.96,2.90/42.15(H$_\epsilon$/C$_\epsilon$)
ILE142	8.14	121.03	3.32/67.00	2.07/37.97(H$_\beta$/C$_\beta$); 1.76,0.34/27.87(H$_{\gamma1}$/C$_{\gamma1}$); 0.45/17.48(H$_{\gamma2}$/C$_{\gamma2}$); 0.07/13.11(H$_\delta$/C$_\delta$)
GLU143	8.21	117.57	3.48/58.75	2.28,1.77/30.49(H$_\beta$/C$_\beta$); 2.18,2.08/36.61(H$_\gamma$/C$_\gamma$)
SER144	7.33	113.60	4.14/60.79	3.75,3.70/62.93(H$_\beta$/C$_\beta$)
PHE145	7.71	119.70	4.96/57.59	3.54,2.97/39.23(H$_\beta$/C$_\beta$); 7.07(H$_\delta$); 7.29(H$_\epsilon$); 7.29(H$_\zeta$)
GLY146	7.80	105.20	4.47,3.61/44.28	-
SER147	8.23	110.03	4.76/57.92	4.15,4.15/66.18(H$_\beta$/C$_\beta$)
SER149	-	-	4.25/-	3.88,3.88/64.58(H$_\beta$/C$_\beta$)
GLY150	8.08	113.39	4.24,3.86/44.48	-
LYS151	7.30	119.70	4.43/56.61	1.72,1.72/32.73(H$_\beta$/C$_\beta$); 1.52,1.52/24.67(H$_\gamma$/C$_\gamma$); 1.74,1.72/(H$_\delta$/C$_\delta$); 3.04,3.04/42.15(H$_\epsilon$/C$_\epsilon$)
THR152	8.76	116.65	5.48/59.82	4.58/71.18(H$_\beta$/C$_\beta$); 1.29/22.34(H$_\gamma$/C$_\gamma$)
SER153	9.40	116.55	4.40/59.04	4.15,4.03/63.51(H$_\beta$/C$_\beta$)
LYS154	7.36	118.38	4.48/54.58	1.16,0.75/38.17(H$_\beta$/C$_\beta$); 1.31,1.01/25.83(H$_\gamma$/C$_\gamma$); 1.22,0.81/(H$_\delta$/C$_\delta$);

Residue	δ(**H**^{N}) [ppm]	δ(^{15}**N**) [ppm]	δ(**H**$_\alpha$/**C**$_\alpha$) [ppm]	**others** [ppm]
				2.90,2.82/41.68(H_ϵ/C_ϵ)
LYS155	8.86	123.57	4.18/56.91	1.75,-/32.05(H_β/C_β); 1.43,1.20/24.96(H_γ/C_γ); 1.64,1.64/(H_δ/C_δ); 2.89,2.89/41.07(H_ϵ/C_ϵ)
ILE156	9.30	132.32	4.94/59.33	2.36/36.90(H_β/C_β); 1.53,1.35/26.03($H_{\gamma1}$/$C_{\gamma1}$); 0.75/19.52($H_{\gamma2}$/$C_{\gamma2}$); 0.69/13.31(H_δ/C_δ)
VAL157	8.88	127.64	4.53/60.89	1.82/36.22(H_β/C_β); 0.74/20.69($H_{\gamma1}$/$C_{\gamma1}$); 0.80/22.53($H_{\gamma2}$/$C_{\gamma2}$)
ILE158	9.10	126.52	4.55/60.89	2.06/37.00(H_β/C_β); 0.90/19.23($H_{\gamma2}$/$C_{\gamma2}$); 0.70/13.70(H_δ/C_δ)
THR159	8.82	123.88	4.00/64.19	4.20/68.07(H_β/C_β); 1.18/22.24(H_γ/C_γ)
ASP160	8.10	116.45	4.93/53.80	3.01,2.92/44.57(H_β/C_β)
CYS161	8.29	115.43	4.48/55.26	3.07,3.01/30.69(H_β/C_β)
GLY162	6.59	101.61	3.82,3.35/45.54	-
GLN163	9.11	120.92	5.03/54.77	1.93,1.64/30.88(H_β/C_β); 2.59,2.45/33.02(H_γ/C_γ); 7.98,7.17/110.54(H_δ/N_δ)
LEU164	9.07	127.23	4.54/55.55	1.64,1.64/43.41(H_β/C_β); 1.62/27.48(H_γ/C_γ); 0.86/26.32($H_{\delta1}$/$C_{\delta1}$); 0.75/23.41($H_{\delta2}$/$C_{\delta2}$)

4.1.2.2 Solution structure of Cyclophilin D

The structure of Cyclophilin D consits of nine anti-parallel β-strands (residue Leu5 to residue Ala12 (β_1), residue Gln15 to residue Leu24(β_2), residue Leu51 to residue Ser52 (β_3), residue Arg55 to residue Ile57(β_4), residue Met61 to residue Ala64 (β_5), residue Val97 to residue Met100 (β_6), residue Phe112 to residue Cys115 (β_7), residue Val127 to residue Glu134 (β_8), and residue Ile156 to residue Gln163 (β_9) forming a β-barrel structure, which exhibits a significant deviation from the ideal cylindrical barrel symmetry. It's best viewed as a sandwich of two roughly orthogonal β-sheets, in which the first sheet is composed of strand β_1 to strand β_3 and strand β_9, and the second sheet is formed by strand β_4 to strand β_7, whereas strand β_8 is part of both β-sheets. The β-barrel is capped by two amphipathic α-helices (residue 30 to residue 41 (α_1) and residue Met136 to residue Ser144 (α_2)) orientated roughly perpendicular to each

other (calculated helix angle is 80,1°). A short $3/10$-helix (residue Asp120 to residue Leu122) inserted in loop β_7/β_8 is located alongside to the β-barrel (see figure 4.11).

Between these regular secondary structure elements, several well defined β-turns of type I (residue Glu43 to residue Gly45, residue 58 to residue 60, residue70 to residue 72, residue 78 to residue 80, and residue 105 to residue 107) and of type II (residue Asp13 to residue Gly15 , residue Lys49 to residue 51, residue 123 to residue 125, and residue 147 to residue 159) together with some β-bulges (residues Leu17-Gly18/Val10, His54-Arg55/Gln63, Phe60-Met61/Ile57, Phe129-Gly130/Leu98, Lys133-Glu134/Val21 and Thr159-Ser160/Asp9) could be identified.

Non surprisingly, the tertiary fold of cyclophilin D basically renders the topology of the highly sequence homologue cyclophilin A published by [133]. One exception is an a β-strand 9 was found to be further stabilized by additional hydrogen bonding to a short β-strand fragment β_3 (residue Ser51 to residue Thr52).

About 31% of the in total 165 residues of wildtype cyclophilin D are involved in β-sheet structure, approximately 12.7% are part of the α-helices, and 1.8% are located in the short $3/10$-helix.

The structural statistics of the final 15 structures for Cyclophilin D are given in table 4.9, RMSD values in table 4.10.

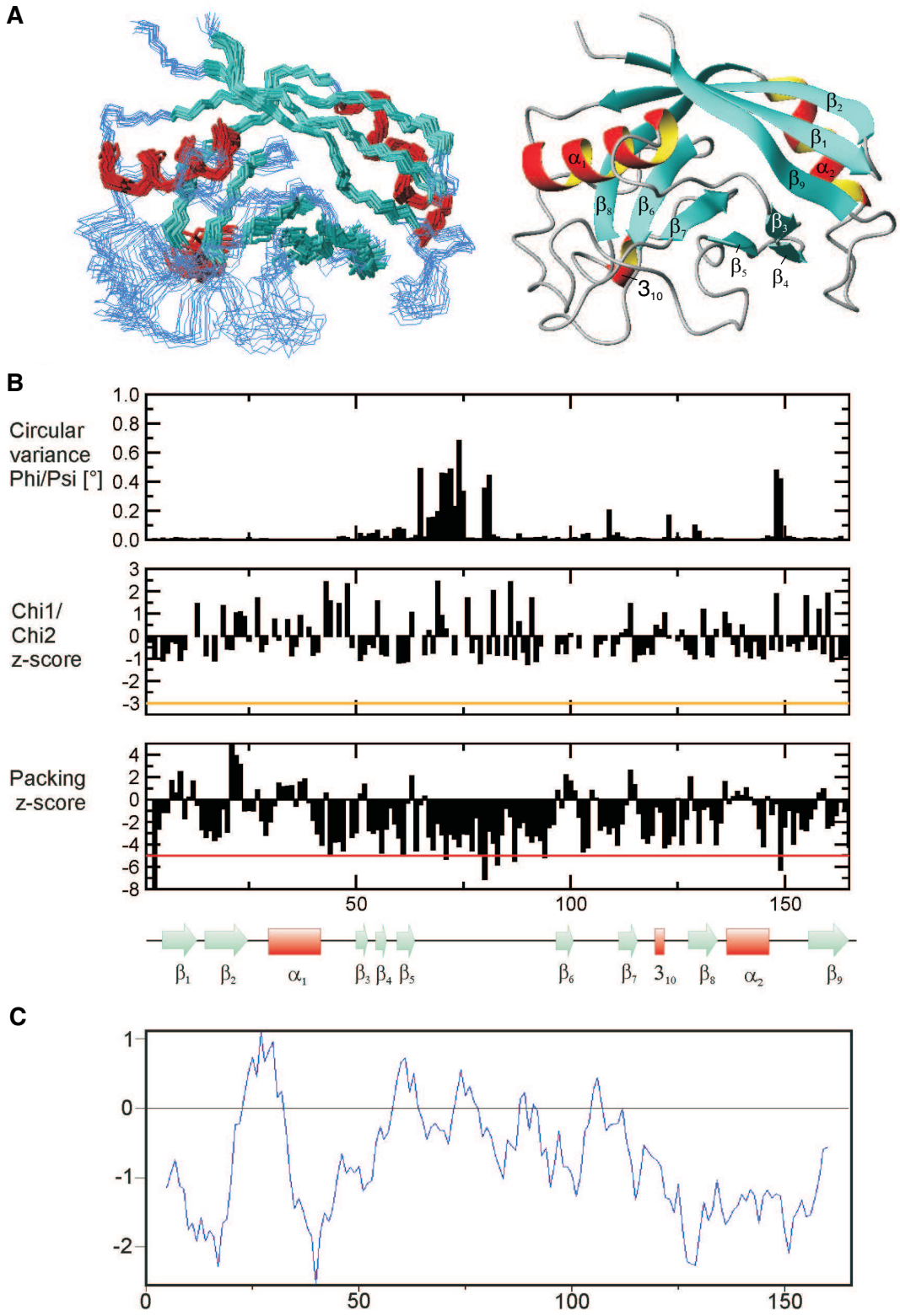


Figure 4.11: (A) Ribbon plot of the backbone structure of recombinant $homosapiens$ cyclophilin D is printed with β-strands in cyan and α-helices in red/yellow beside a representative ensemble of 15 models. In (B) the precision of the bundle in term of circular variances, WHATif z-score values for sidechain $\chi 1/\chi 2$ torsions and the packing quality per residue is plotted. At bottom the secondary structure elements are given. (C) PROSA total interaction energy $\Delta S(S, C)$ profi le applying a window size of 10 residues.

Table 4.9: Structural statistics for the fi nal 15 Cyclophilin D structures.

Experimental NOE distance restraints	
total NOE distance restraints	1883
intra-residue distance restraints	518
sequential backbone ($\|i-j\|$=1) distance restraints	556
inter-residue, medium-range (4$\leqq |i-j| \leqq$2) distance restraints	204
inter-residue, long-range ($|i-j| >$4) distance restraints	605
Torsions restraints	
backbone-ϕ,ψ-torsions (TALOS)	252
Hydrogen bonds	
N-H$\cdots$O=C	58
Average RMS-deviations from	
experimental distance restraints [Å]	0.042
idealized covalent geometry	
bonds[Å]	4.980E-03 $\pm$ 1.071E-04
angles[$^\circ$]	6.946E-01 $\pm$ 1.334E-02
improper[$^\circ$]	4.759E-01 $\pm$ 1.264E-02
PROCHECK statistics	
Percentage residues in allowed regions of Ramachandran plot	
% in whole molecule (% in ordered structure elements[a])	
most favoured regions [A,B,L]	92.1%
additional allowed regions [a,b,l,p]	6.2%
generously allowed regions [$\sim$a, $\sim$b, $\sim$l, $\sim$p]	1.3%
total in allowed regions	99.7% (100%)
Omega angle standard deviation (typical value[b])	0.9 (6.0)
Zeta angle standard deviation (typical value[b])	0.6 (3.1)
H-bond energy standard deviation (typical value[b])	0.7 (0.8)
PROSA	
Mean force potential - whole molecule (residues in ordered structure elements[a])	
pair energy	-0.34 (-0.46)
surface energy	-0.45 (-0.59)
combined energy	-0.79 (-1.05)
z-score - whole molecule (omitting unstructured N- and C-terminus)	
pair energy	-4.24
surface energy	-6.02
combined energy	-7.24
What if	
Overall packing score including residues	
2-165	-1.71

Table 4.10: Root mean square deviations to the mean structure of Cyclophilin D.

Selected residues	Selected atoms	RMSD [Å]
all elements of secondary structure	backbone atoms	0.54
	all heavy atoms	1.04
	all atoms	1.34
all β-strands	backbone atoms	0.52
	all heavy atoms	1.05
	all atoms	1.35

[a] statistic apply to ordered regions of cyclophilin D, with residues 5-12 (β_1), 15-24 (β_2), 30-41 (α_1) 51-52 (β_3), 55-57 (β_4), 61-64 (β_5), 97-100 (β_6), 112-115 (β_7), 120-122 ($3/10$-helix), 127-133(β_8),136-145(α_2),156-164(β_9) [b] according [107]

4.1.2.3 Comparison of the NMR solution and the crystal structures of cyclophilin D

Because of the high isoelectric point of about 10.2 and the extreme solubility of Cyclophilin D the wild type protein was not suitable for crystallization. However, the K133I mutant [160] gave crystals diffracting to 1.5Å. The resulting high resolution structure was taken here for comparison. The global RMSD calculated for all backbone N, C^{α}, and C' atoms (residue Gln3 to residue Gln163) is 1.93Å. Taking solely atoms of regularly secondary structure and the β-barrel into account RMSD values of 1.36 and 1.18 are obtained, respectively. Beside the C- and N-termini the most significant backbone differences concerned residues of loop β_4/β_5 (residue Pro58 to residue Phe60), loop β_5/β_6 (in particular residue Asn69 to residue Gly74 and residue Gly80 to Ser81), loop β_6/β_7 (most pronounced Gly104 to residue Asn108), and loop α_2/β_9 (particularly residue S147 to residue Lys151). In the cyclophilin A homologue the mentioned loops are involved in cyclosporin A binding (figure 4.12.B, top). At the NMR solution structure of cyclophilin D these loops are conspicuously more disordered and appear to be more displaced towards the free space of the homologue binding pocket. Moreover, minor differences concern the additional existence of a two residue strand β_3 (supplementary stabilizing strand β_9, colored in green in figure 4.12.A) and to more shorten strand β_4 (strand β_3 in 2BIT.pdb) compared to the x-ray structure as explicitly indicated from the NOE pattern and secondary chemical shift analysis. At the deposited models of the homologous cyclophilin A (1CNA, NMR model) and cyclophilin B (1CYN, x-ray structure) that two residue part (residue Ser51 and β-branched residue Thr52 in β_3 strand of the cycD NMR model, Ser51-Cys52 in cycA, and Ser69-Lys70 in cycB) is found conspicuously spatial distant from strand β_9.

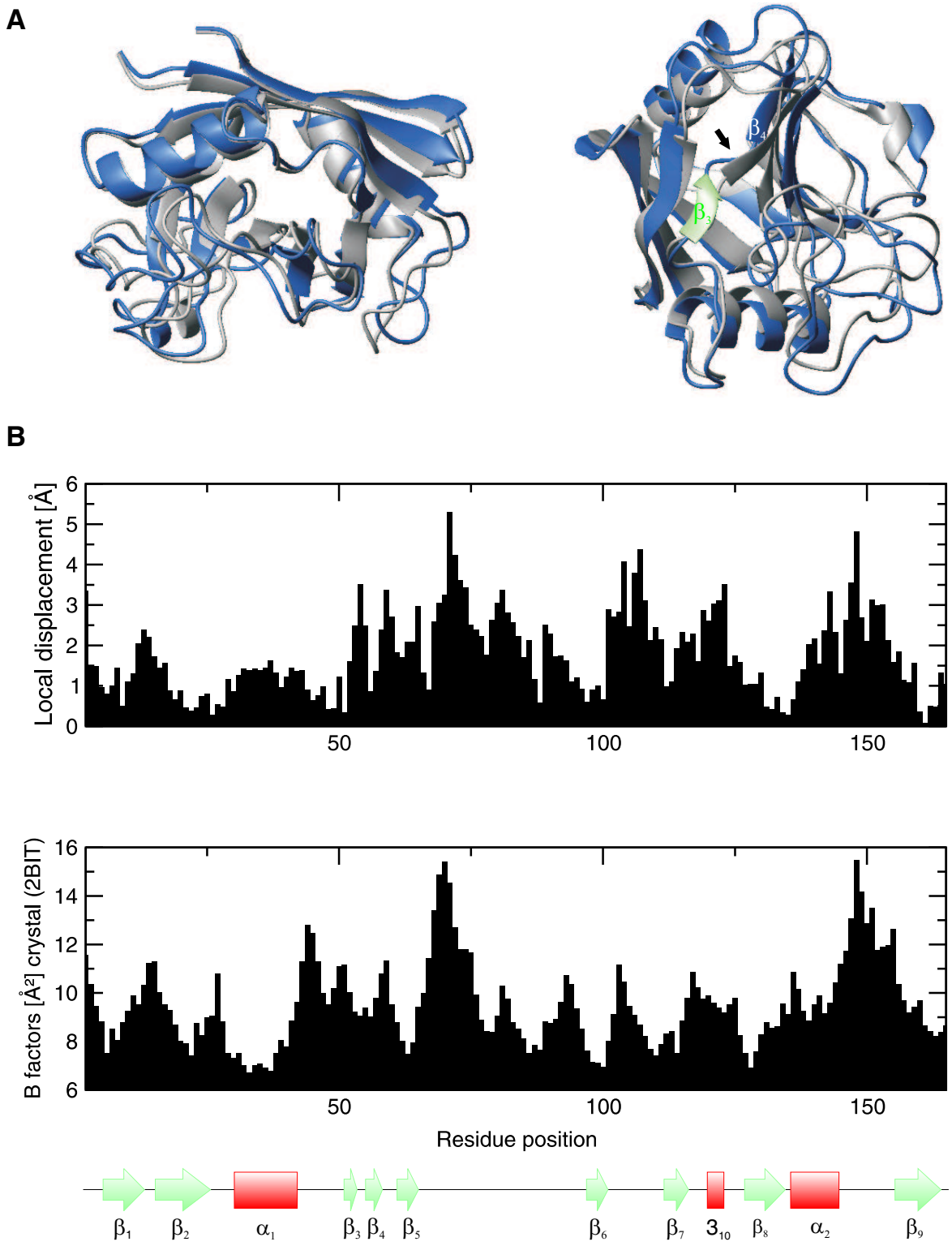


Figure 4.12: (A) Overlay with the NMR and x-ray (2BIT.pdb; [133]) structure of cyclophilin D. The arrow indicate the position of the more shorten strand β_4 (strand β_3 in 2BIT.pdb), strand β_3 of the NMR model is colored in green. (B) Global displacements per residue of both structures (top), and temperature B factors per residue for backbone C^{α} atoms in 2BIT (middle). At bottom the secondary structure elements of cyclophilin D as inferred from $\Delta^{13}C^{\alpha}$-$\Delta^{13}C^{\beta}$ secondary chemical shifts and NOE pattern analysis are shown.

4.1.3 LIM1

4.1.3.1 Chemical shift assignment

High quality spectra with an excellent $^1H^N$ and ^{15}N chemical shift dispersion were obtained for *quail* LIM1 domain of CRP2. Only two signal pairs were found almost exactly at the same resonance position (residue His31 and residue Cys40, and residue Tyr18 and residue Met38, respectively). The backbone chemical shift assignment was initially performed by Elisabeth Dowler taking the published backbone amide shifts [84] as reference. In total sequential backbone resonance assignment were derived for 72 (of in total 79) non-proline residues. In addition to the N-terminal methionine, assignment was not possible for Trp4, Gly5-7, Ser46, and Lys69. By dint of the uniformly ^{13}C,^{15}N labeled sample used in this work 82.7% sidechain ^1H,^{13}C aliphatic- and 65.9% ^1H aromatic chemical shifts could be assigned. Sidechain function for which an assignment was not possible were located either at the unstructured N-terminal (residue Met1 to residue Gly7) or C-terminal (residue Gly67 to residue Asn81) part of the CRP2 LIM1 domain construct. Furthermore, ^1H,^{15}N chemical shift assignment for 3 (out of 6 observable) sidechain amide functions (all resident in the folded part of LIM1) was possible (for 2 asparagine δ-NH_2 and for 1 ϵ-NH_2 glutamine groups).

Table 4.11: Backbone and sidechain chemical shifts of *quail* LIM1(CRP2).

residue	δ(**H**^{N}) [ppm]	δ(^{15}**N**) [ppm]	δ(**H**$_\alpha$/**C**$_\alpha$) [ppm]	**others** [ppm]
PRO2	-	-	4.16/62.62	2.13,1.75/32.00(H_β/C_β); 25.12(C_γ);
ASN3	8.11	116.13	4.77/-	-
TRP4	-	-	-/-	-
GLY5	-	-	-/-	-
GLY6	-	-	-/-	-
GLY7	-	-	3.85/-	-
ASN8	8.32	119.10	4.72/50.97	2.55,2.49/38.22(H_β/C_β); 7.8,6.80/114.44(H_δ/N_δ)
LYS9	8.28	119.95	4.04/53.17	1.51,1.29/31.81(H_β/C_β); 1.26,1.15/22.87(H_γ/C_γ); 1.49,1.49/(H_δ/C_δ); 2.91,2.93/-(H_ϵ/C_ϵ)
CYS10	8.24	125.46	4.03/56.47	3.41,2.31/29.67(H_β/C_β)
GLY11	8.69	116.35	3.81,3.16/45.40	-
ALA12	9.08	123.98	5.02/52.00	1.83/19.57 (H_β/C_β)
CYS13	8.17	115.50	5.03/57.44	3.36,3.07/30.45(H_β/C_β)
GLY14	8.03	111.68	4.16,3.80/44.82	-

residue	δ(**H**^{N}) [ppm]	δ(**^{15}N**) [ppm]	δ(**H$_\alpha$/C$_\alpha$**) [ppm]	**others** [ppm]
ARG15	8.40	120.80	4.74/53.82	2.10,2.01/30.25(H$_\beta$/C$_\beta$); 1.81,1.71/26.37(H$_\gamma$/C$_\gamma$); 3.26,3.28/(H$_\delta$/C$_\delta$)
THR16	8.64	118.25	3.95/63.46	3.76/67.54(H$_\beta$/C$_\beta$); 0.70/20.35(H$_\gamma$/C$_\gamma$)
VAL17	8.66	128.01	4.02/58.02	1.15/31.42(H$_\beta$/C$_\beta$); 0.48/19.38(H$_{\gamma1}$/C$_{\gamma1}$); 0.15/18.79(H$_{\gamma2}$/C$_{\gamma2}$)
TYR18	9.09	130.31	4.69/56.28	3.23,2.87/37.05(H$_\beta$/C$_\beta$); 7.22(H$_\delta$); 6.79(H$_\epsilon$)
HIS19	8.59	121.01	4.24/58.41	3.20,3.09/28.12(H$_\beta$/C$_\beta$)
ALA20	8.02	118.89	4.12/52.78	1.43/16.85 (H$_\beta$/C$_\beta$)
GLU21	7.77	114.44	4.61/53.56	2.53,2.30/29.48(H$_\beta$/C$_\beta$); 2.46,2.35/34.91(H$_\gamma$/C$_\gamma$)
GLU22	7.17	121.01	4.29/55.50	2.04,1.83/30.64(H$_\beta$/C$_\beta$); 1.99,1.84/35.11(H$_\gamma$/C$_\gamma$)
VAL23	9.00	125.67	4.20/59.58	1.59/32.78(H$_\beta$/C$_\beta$); 0.65/18.99(H$_{\gamma1}$/C$_{\gamma1}$); 0.32/18.41(H$_{\gamma2}$/C$_{\gamma2}$)
GLN24	8.45	124.19	5.24/52.78	1.86,1.83/29.28(H$_\beta$/C$_\beta$); 2.17,2.09/32.19(H$_\gamma$/C$_\gamma$); 7.33,6.58/110.83(H$_\delta$/N$_\delta$)
CYS25	8.91	122.92	4.48/56.86	2.59,2.59/27.15(H$_\beta$/C$_\beta$)
ASP26	9.43	130.34	4.36/54.14	3.05,2.56/38.60(H$_\beta$/C$_\beta$)
GLY27	8.66	103.62	4.11,3.62/44.02	-
ARG28	7.92	120.80	4.47/53.25	1.88,1.55/31.42(H$_\beta$/C$_\beta$); 1.63,1.58/25.59(H$_\gamma$/C$_\gamma$); 3.20,3.16/(H$_\delta$/C$_\delta$)
SER29	8.39	116.98	4.91/57.25	3.40,3.23/63.27(H$_\beta$/C$_\beta$)
PHE30	8.86	118.25	5.62/54.43	3.05,2.30/41.90(H$_\beta$/C$_\beta$); 7.02(H$_\delta$); 7.17(H$_\epsilon$); -($_\zeta$)
HIS31	8.63	119.74	4.68/57.79	3.77,3.73/29.75(H$_\beta$/C$_\beta$)
ARG32	9.37	127.26	3.99/59.44	1.98,1.93/27.73(H$_\beta$/C$_\beta$); 1.70,1.67/26.37(H$_\gamma$/C$_\gamma$); 3.33,3.23/(H$_\delta$/C$_\delta$)
CYS33	8.93	113.16	4.48/58.22	3.23,2.92/25.59(H$_\beta$/C$_\beta$)
CYS34	7.65	118.47	4.42/58.99	3.48,3.21/29.86(H$_\beta$/C$_\beta$)
PHE35	7.60	126.52	4.43/54.72	3.34,2.61/35.69(H$_\beta$/C$_\beta$); 6.52(H$_\delta$); 6.82(H$_\epsilon$); 6.82($_\zeta$)
LEU36	7.21	122.07	4.78/50.75	1.15,1.05/43.85(H$_\beta$/C$_\beta$); 1.21/25.40(H$_\gamma$/C$_\gamma$); 0.73/23.65(H$_{\delta1}$/C$_{\delta1}$); 0.70/22.87(H$_{\delta2}$/C$_{\delta2}$)
CYS37	8.22	121.65	3.88/58.44	3.31,2.75/29.48(H$_\beta$/C$_\beta$)
MET38 MET38	9.09	130.13	4.07/57.44	2.60,2.45/31.03(H$_\beta$/C$_\beta$); 2.04,1.99/31.03(H$_\gamma$/C$_\gamma$);

residue	δ(**H**^{N}) [ppm]	δ(**^{15}N**) [ppm]	δ(**H$_\alpha$/C$_\alpha$**) [ppm]	**others** [ppm]
				0.69/18.60(H$_\epsilon$/C$_\epsilon$)
VAL39	9.15	122.07	4.12/63.66	2.80/31.03(H$_\beta$/C$_\beta$); 1.00/19.96(H$_{\gamma1}$/C$_{\gamma1}$); 1.18/20.74(H$_{\gamma2}$/C$_{\gamma2}$)
CYS40	8.63	119.74	4.78/58.46	3.28,2.68/30.46(H$_\beta$/C$_\beta$)
ARG41	7.85	116.13	4.14/55.50	2.13,2.09/24.23(H$_\beta$/C$_\beta$); 1.58,1.44/25.01(H$_\gamma$/C$_\gamma$); 3.11,3.11/(H$_\delta$/C$_\delta$)
LYS42	8.15	120.27	4.25/55.69	1.99,1.65/32.00(H$_\beta$/C$_\beta$); 1.49,1.45/23.07(H$_\gamma$/C$_\gamma$); 1.73,1.73/(H$_\delta$/C$_\delta$); 3.08,3.08/40.35(H$_\epsilon$/C$_\epsilon$)
ASN43	-	-	4.50/53.17	2.83,2.63/37.24(H$_\beta$/C$_\beta$); 7.5,6.97/112.53(H$_\delta$/N$_\delta$)
LEU44	7.67	123.13	4.28/51.99	0.51,0.42/42.23(H$_\beta$/C$_\beta$); 1.23/24.23(H$_\gamma$/C$_\gamma$); -0.17/24.81(H$_{\delta1}$/C$_{\delta1}$); 0.44/20.93(H$_{\delta2}$/C$_{\delta2}$)
ASP45	7.61	116.56	4.85/51.42	3.18,2.79/41.52(H$_\beta$/C$_\beta$)
THR47	8.84	112.74	4.56/61.52	4.37/67.93(H$_\beta$/C$_\beta$); 1.26/20.35(H$_\gamma$/C$_\gamma$)
THR48	7.75	111.26	4.48/59.77	4.44/69.52(H$_\beta$/C$_\beta$); 1.27/21.13(H$_\gamma$/C$_\gamma$)
VAL49	7.42	120.80	4.39/59.77	1.83/30.83(H$_\beta$/C$_\beta$); 1.00/18.21(H$_{\gamma1}$/C$_{\gamma1}$); 0.77/21.32(H$_{\gamma2}$/C$_{\gamma2}$)
ALA50	8.97	130.98	4.33/49.87	0.69/19.57 (H$_\beta$/C$_\beta$)
ILE51	8.18	121.65	4.91/58.41	1.73/38.60(H$_\beta$/C$_\beta$); 1.52,1.07/26.76(H$_{\gamma1}$/C$_{\gamma1}$); 0.91/16.27(H$_{\gamma2}$/C$_{\gamma2}$); 0.86/12.00(H$_\delta$/C$_\delta$)
HIS52	8.91	126.73	4.77/54.48	3.35,2.72/30.25(H$_\beta$/C$_\beta$)
ASP53	9.01	127.79	4.00/55.30	2.88,2.73/37.24(H$_\beta$/C$_\beta$)
ALA54	8.32	124.61	4.52/51.03	1.54/17.44 (H$_\beta$/C$_\beta$)
GLU55	8.50	119.95	4.40/53.95	2.70,2.06/30.64(H$_\beta$/C$_\beta$); 2.58,2.38/34.33(H$_\gamma$/C$_\gamma$)
VAL56	7.78	119.53	4.34/59.58	1.72/32.00(H$_\beta$/C$_\beta$); 0.44/20.35(H$_{\gamma1}$/C$_{\gamma1}$); 0.69/18.41(H$_{\gamma2}$/C$_{\gamma2}$)
TYR57	8.35	121.22	5.83/54.72	3.74,2.93/43.26(H$_\beta$/C$_\beta$); 7.18(H$_\delta$); 6.91(H$_\epsilon$)
CYS58	9.50	116.35	5.45/55.89	3.41,2.82/30.25(H$_\beta$/C$_\beta$)
LYS59 LYS59	8.40	119.95	3.78/59.01	1.86,1.82/30.84(H$_\beta$/C$_\beta$); 1.44,1.28/25.01(H$_\gamma$/C$_\gamma$); 1.66,1.63/(H$_\delta$/C$_\delta$); 2.96,2.92/37.44(H$_\epsilon$/C$_\epsilon$)

residue	δ(**H**^{N}) [ppm]	δ(**^{15}N**) [ppm]	δ(**H$_\alpha$/C$_\alpha$**) [ppm]	**others** [ppm]
SER60	8.26	115.50	4.34/60.16	4.14,4.12/61.13(H$_\beta$/C$_\beta$)
CYS61	9.37	126.73	3.93/63.46	3.23,3.02/28.12(H$_\beta$/C$_\beta$)
TYR62	8.85	120.37	3.35/60.35	2.96,2.91/37.44(H$_\beta$/C$_\beta$); 7.23(H$_\delta$); 6.75(H$_\epsilon$)
GLY63	8.16	104.90	3.90,3.72/45.01	-
LYS64	7.35	119.53	3.99/56.86	1.83,1.83/31.03(H$_\beta$/C$_\beta$); 1.55,1.40/23.84(H$_\gamma$/C$_\gamma$); 1.66,1.59/(H$_\delta$/C$_\delta$); 2.90,2.90/40.55(H$_\epsilon$/C$_\epsilon$)
LYS65	7.27	118.47	3.85/56.47	1.16,0.83/32.00(H$_\beta$/C$_\beta$); 0.11,0.05/22.87(H$_\gamma$/C$_\gamma$); 1.29,1.09/(H$_\delta$/C$_\delta$); 2.74,2.68/40.35(H$_\epsilon$/C$_\epsilon$)
TYR66	8.11	114.86	4.42/56.28	2.80,1.56/37.24(H$_\beta$/C$_\beta$); 6.30(H$_\delta$); 6.30(H$_\epsilon$)
GLY67	7.49	108.71	4.15,4.10/43.26	-
GLY70	8.22	108.71	-/43.87	-
TYR71	7.88	119.74	-/56.65	37.18(C$_\beta$);
GLY72	7.54	109.14	3.87,3.87/43.64	-
TYR73	7.98	119.53	-/56.79	2.92/36.96(H$_\beta$/C$_\beta$);
GLY74	8.43	110.73	3.90/43.87	-
GLN75	8.24	120.01	4.36/54.43	27.78(C$_\beta$); 32.62(C$_\gamma$);
GLY76	8.25	109.35	-,-/43.87	-
ALA77	8.28	124.06	-/-	-
GLY78	8.48	108.76	3.90/43.87	-
THR79	8.03	113.80	-/-	-
LEU80	8.38	124.88	-/-	-

As shown in table 4.12, a fourfold number of long-range NOE compared to the input data of the originally published *quail* LIM1 domain structure of CRP2 could be derived from the 3D ^{13}C- and ^{15}N edited NOESY spectra for the fold refinement resulting in a final model ensemble of high precision (see next section). A plot of the secondary shift is given in figure 4.13.

4.1.3.2 Solution structure of the LIM1 domain

The tertiary structure of *quail* CRP2-LIM1 includes rather short regions of regular secondary structure connected by flexible loop regions. Both zinc coordinating subdomains display a similar fold containing two anti-parallel β-sheets. The two zinc fingers subdomains pack together forming a hydrophobic interface. The N-terminal Cys-Cys-His-Cys (CCHC) zinc finger subdomain I (residue 8-33) starts with an anti-parallel sheet (strand β_1 and strand β_2; residue Lys9 to residue Cys10 and residue Thr16 to residue Tyr18, respectively) connected by a rubredoxin-

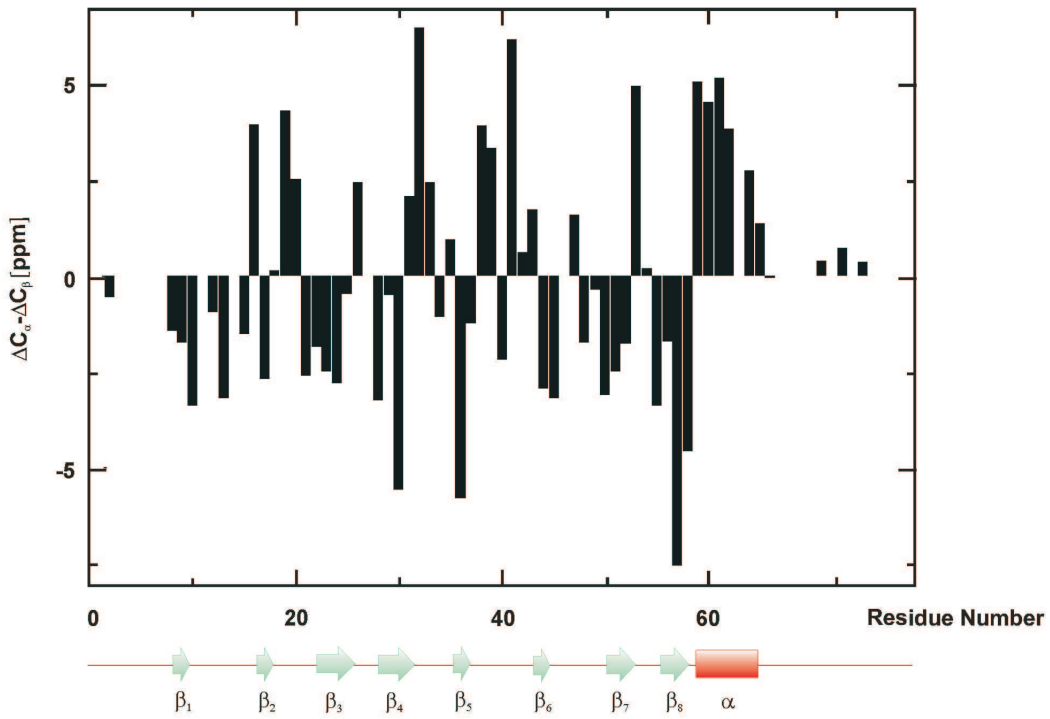


Figure 4.13: $\Delta^{13}C^{\alpha}$-$\Delta^{13}C^{\beta}$ secondary chemical shifts of $quail$ LIM1/CRP2 as a function of the residue position.

type turn that accommodates the two Zn^{2+} coordinating residues Cys10 and Cys13 of zinc finger I. Hydrogen bonds could be inferred from NOE pattern between the backbone amide H^N protons of Ala 12 or Arg15 and sidechain S^{γ} of Cys10 or Cys13, respectively, a characteristic finding at the CCHC site in LIM domains [83]. A second anti-parallel β-sheet (strand β_3 and strand β_4; residue Val23 to residue Cys25 and Arg28 to residue His31, respectively) accommodates the third (His31) and a subsequent linker region, connecting both subdomains I and II, the fourth (Cys33) zinc-coordinating site of domain I.

The following third anti-parallel β-sheet belonging to subdomian II is roughly orientated perpendicular the second one, whereas its constituting β-strands (strand β_5 and strand β_6; residue Phe35 to residue Cys37 and residue Lys42 to residue Leu44, respectively) are also connected by a rubredoxin-type turn including the Zn^{2+} coordinating residues Cys37 and Cys40 of domain II. The fourth anti-parallel β-sheet (strand β_7 and strand β_8; residue Ala50 to residue His52 and residue Glu55 to residue Cys58) contains the third (Cys58) and the C-terminal α helix the last (Cys61) zinc-coordinating site of domain I. The helical stretch ranges from residue Cys61 to residue Gly67.

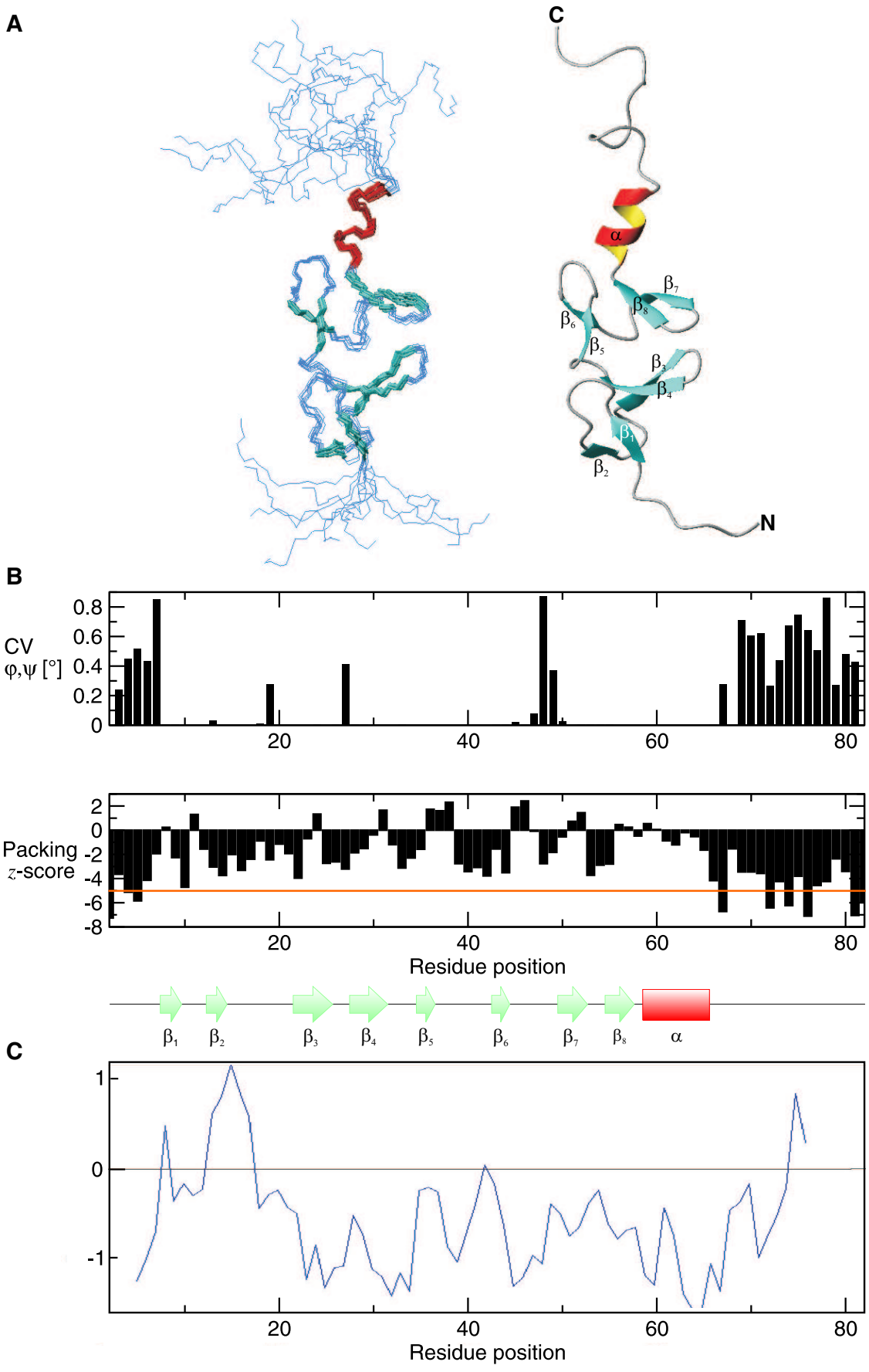


Figure 4.14: (A) Ribbon plot of the backbone structure of N-terminal recombinant $quail$ LIM1 is drawn with β-strands in cyan and α-helices in red/yellow right to a representative ensemble of 15 models. In (B) the precision of the bundle in term of circular variances and the packing quality per residue is plotted.

Thus, the Cys-Cys-Cys-Cys (CCCC) zinc finger of subdomain II (residue 34 to residue 67) exhibits an essentially identical fold appearance to the CCHC zinc finger of subdomain I. Analogically, characteristic hydrogen bonding pattern was found between backbone amide H^N of residue Val39 or residue Lys42 to S^{gamma} of residue Cys37 or residue Cys40.

The structural statistics of final 15 structures of *quail* CRP2-LIM1 are summarized in table 4.12.

Table 4.12: Structural statistics for the best fi nal 15 *quail* LIM1/CRP2 structures.

Experimental NOE distance restraints	
total NOE distance restraints	1332 (468)
intra-residue distance restraints	458 (158)
sequential backbone ($\|i-j\|$=1) distance restraints	303 (143)
inter-residue, medium-range (4$\leqq |i-j| \leqq$2) distance restraints	158 (55)
inter-residue, long-range ($|i-j| >$4) distance restraints: 413 (112)	
Torsions restraints	
backbone-ϕ,ψ-torsions (TALOS)	22
Hydrogen bonds	
N-H$\cdots$O=C	16
N-H$\cdots S^{\gamma}$	4
Average RMS-deviations from	
experimental distance restraints [Å]	0.042
idealized covalent geometry	
bonds[Å]	4.980E-03 $\pm$ 1.071E-04
angles[°]	6.946E-01 $\pm$ 1.334E-02
improper[°]	4.759E-01 $\pm$ 1.264E-02
PROCHECK statistics	
Percentage residues in allowed regions of Ramachandran plot	
% in whole molecule (% in ordered structure elements[a])	
most favoured regions [A,B,L]	87.2% (96.3%)
additional allowed regions [a,b,l,p]	?% (3.7%)
generously allowed regions [$\sim$a, $\sim$b, $\sim$l, $\sim$p]	?% (0.0%)
total in allowed regions	97.1% (100%)
Omega angle standard deviation (typical value[b])	1.3 (6.0)
Zeta angle standard deviation (typical value[b])	1.1 (3.1)
H-bond energy standard deviation (typical value[b])	0.9 (0.8)
PROSA	
Mean force potential - whole molecule (residues in ordered structure elements[a])	
pair energy	-0.50 (-0.44)
surface energy	-0.10 (-0.24)
combined energy	-0.60 (-0.68)
z-score - whole molecule (omitting unstructured N- and C-terminus)	
pair energy	-4.00
surface energy	-3.06

Experimental NOE distance restraints	
combined energy	-4.10
What if	
Overall packing score including residues	
8-66 (omitting unstructured N- and C-terminus)	-1.28

Approximately, a threefold number of experimental NOE distance restraints compared to input data of the prior published model [84] could be defined resulting in a suitable refined structure model for diffusion tensor analysis (subsequent performed by Paul Schanda and Elizabeth Dowler, unpublished). As indicated by the RMSD values (see table), subdomain II appears to be slightly more compactly defined. The overall (all atoms) and backbone RMS deviation for the final structure ensemble (15 models) was 0.45Å and 1.17Å, respectively, compared to 1.30Å and 2.02Å of the initially published solution models, respectively.

Table 4.13: Root mean square deviations to the mean structure of LIM1

Whole LIM1 domain (residue 8 to residue 67)		
Selected residues	**Selected atoms**	**RMSD [Å]**
all elements of secondary structure	backbone atoms	0.45 (1.30)
	all heavy atoms	0.95
	all atoms	1.17 (2.02)
all β-strands	backbone atoms	0.35
	all heavy atoms	0.66
	all atoms	0.90
N-terminal Cys-Cys-His-Cys zinc finger subdomain (residue 8 to residue 32)		
Selected residues	**Selected atoms**	**RMSD [Å]**
all elements of secondary structure	backbone atoms	0.25 (0.72)
	all heavy atoms	0.65
	all atoms	0.90 (1.78)
C-terminal Cys-Cys-Cys-Cys zinc finger subdomain (residue 33 to residue 67)		
Selected residues	**Selected atoms**	**RMSD [Å]**
all elements of secondary structure	backbone atoms	0.35 (0.83)
	all heavy atoms	1.00
	all atoms	1.22 (1.78)
all β-strands	backbone atoms	0.22
	all heavy atoms	0.54
	all atoms	0.80

[a] statistic apply to ordered regions of LIM1, with residues 8-9 ($\beta_{1)}$),16-17 ($\beta_{2)}$),22-25 (β_3),28-31 (β_4),35-36 (β_5),43-44 (β_6),50-52 (β_7),55-57 (β_8), and 59-66 (C-terminal α-helix). [b] according [107].

4.2 Resonance assignment and relaxation of intrinsic unfolded proteins

4.2.1 BASP1

The $^{1}H^{N}$-^{15}N HSQC (see figure 4.15) show a distinct separation of the cross peaks into regions typical of residue random coil positions. There are perspicuously apparent resonance clusters for glycines, serines, and threonines at the top and alanines at the bottom of the spectrum. Furthermore, residue types with similar sequential neighborhood appear narrowly at the same place in the $^{1}H^{N}$-^{15}N chemical shift plane, which provided a 'cross-validation' check feature in the manual assignment process.

However, one underlying problem in resonance assignment of an intrinsic unfolded protein, such as BASP1, is the poor extent of backbone amide $^{1}H^{N}$ chemical shift dispersion (about 0.6 ppm for BASP1) compared to a protein molecule at folded state. In addition, because of the almost entirely absence of secondary chemical shifts the $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ resonances were found preponderantly at random coil chemical shift values, leading to hard signal overgrowing along the carbon chemical shift axis. On the other side, once a spinsystem is unambiguously defined, its related residue type can be generally determined by dint of the tabulated random coil shifts in a straightforward manner, thus providing an additional assignment tool for allocation of the backbone chemical shifts in an unfolded polypeptide molecule. However, especially owing to the size of BASP1 (245 residues) and therefore extended spectral overlapping unambiguous composing of the inter- and intra backbone chemical shifts for the most spinsystems could be exclusively performed by taking all possible chemical shift combinations of the sequential neighboring spinsystems into account. Note that this could only be done by combination use of *both*, the 3D HNN and the 3D HN(C)N experiment [136], which enable the implementation of the more sequence depending $^{1}H^{N}$ and ^{15}N chemical shifts (beside the inter- and intra carbonyl $^{13}C'$ resonances extracted from the 3D HNCO and 3D HN(CA)CO experiments) for connecting the spinsystems.

Both groups of backbone chemical shifts ($^{15}N_{i-1}/^{15}N_{i}/^{15}N_{i+1}$ and $^{13}C'_{i-1}/^{13}C'_{i}$) in the individual spinsystems permit a high degree of discrimination between them and, accordingly, a large extent of unambiguous backbone resonance assignment.

Because of the large relaxation times inherent to high dynamical unfolded proteins the sensitivity obtained in the triple resonance experiments for BASP1 was excellent.

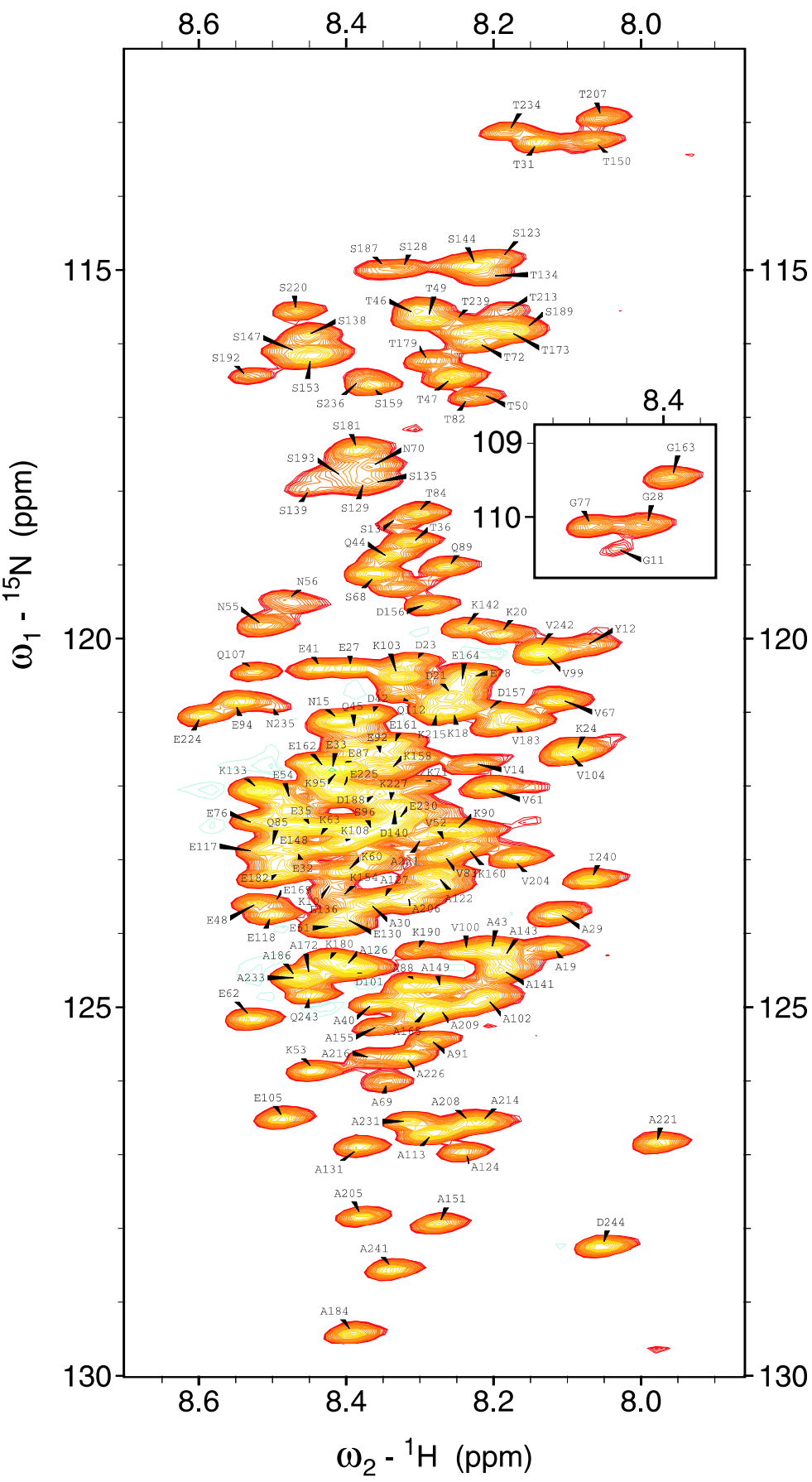


Figure 4.15: 800 MHz 2D $^1H^N$-^{15}N HSQC of *chicken* brain abundant protein 1 (BASP1).

In total 158 of 223 non-proline $^{1}H^{N}$ and ^{15}N backbone chemical shifts were assigned (70.9%). $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ chemical shifts were allocated for 190 residues (77.9%). Peak positions of assigned carbonyl $^{13}C'$ resonances could be unambiguous identified for 172 out of 244 residues (70.5%). The assignment were deposited at the BioMolResBank (BMRB, http://www.bmrb.wisc.edu) entry code 16003 and are given in table 4.14.

Table 4.14: Backbone chemical shift assignment of brain abundant protein 1 (BASP1)

Residue	δ(**$^{1}H^{N}$**) [ppm]	δ(**^{15}N**) [ppm]	δ(**$^{13}C'$**) [ppm]	δ(**$^{13}C^{\alpha}$**) [ppm]	δ(**$^{13}C^{\beta}$**) [ppm]
Lys10	8.42	123.3	176.9	56.6	33.2
Gly11	8.46	110.6	173.6	45.1	-
Tyr12	8.05	120.1	175.7	57.9	39.3
Ser13	8.31	118.5	174.4	57.8	64.2
Val14	8.23	121.8	176.0	62.7	33.0
Asn15	8.40	121.0	175.0	53.3	38.9
Lys18	8.25	121.0	176.8	56.6	32.8
Ala19	8.10	124.2	178.0	52.9	19.2
Lys20	8.18	120.0	176.7	56.7	32.9
Asp21	8.26	120.8	-	54.7	41.2
Asp23	8.29	120.5	176.3	54.6	41.1
Lys24	8.07	121.6	176.8	56.6	32.9
Glu27	8.37	120.5	177.1	56.9	30.4
Gly28	8.41	110.1	173.8	45.3	-
Ala29	8.10	123.7	177.5	52.4	19.4
Ala30	8.36	123.6	178.0	52.5	19.3
Thr31	8.13	113.3	174.6	61.8	69.9
Glu32	8.45	122.9	176.5	56.6	30.4
Glu33	8.40	121.8	-	56.7	30.4
Glu35	8.44	122.6	176.4	56.5	30.1
Thr36	8.31	118.8	172.7	60.0	69.8
Ala40	8.35	125.1	177.9	52.7	19.4
Glu41	8.42	120.4	176.5	56.9	30.4
Asp42	8.35	121.2	176.2	54.4	41.1
Ala43	8.18	124.3	177.9	52.8	19.2
Gln44	8.33	119.0	176.1	56.0	29.5
Gln45	8.37	121.3	176.6	56.0	29.6
Thr46	8.29	115.7	174.6	61.8	69.8
Thr47	8.24	116.5	174.5	62.0	69.8
Glu48	8.50	123.7	176.6	56.7	29.5
Thr49	8.28	115.7	174.7	61.8	69.7
Thr50	8.22	116.6	174.3	62.0	69.7
Glu51	8.41	123.9	176.2	56.4	30.3
Val52	8.26	122.8	176.2	62.4	32.8
Lys53	8.44	125.9	176.5	56.2	33.1

Residue	δ(**^{1}H^{N}**) [ppm]	δ(**^{15}N**) [ppm]	δ(**^{13}C$^{'}$**) [ppm]	δ(**^{13}C$^{\alpha}$**) [ppm]	δ(**^{13}C$^{\beta}$**) [ppm]
Glu54	8.48	122.2	176.2	56.5	30.2
Asn55	8.49	119.8	174.9	53.3	38.9
Asn56	8.46	119.3	175.3	53.4	38.9
Lys60	8.38	123.1	176.3	56.2	33.1
Val61	8.20	122.1	176.2	62.4	33.0
Glu62	8.51	125.2	176.5	56.5	30.3
Lys63	8.40	122.8	-	56.4	33.1
Val67	8.11	120.9	176.4	62.6	32.9
Ser68	8.36	119.2	174.4	58.4	63.8
Ala69	8.34	126.0	179.4	52.6	19.4
Asn70	8.34	117.7	175.2	53.3	38.9
Lys71	8.28	121.9	176.7	56.6	33.1
Thr72	8.22	116.0	174.5	62.2	69.8
Glu76	8.51	122.4	177.1	57.1	30.2
Gly77	8.48	110.2	174.2	45.4	-
Glu78	8.22	120.7	-	56.6	30.4
Thr82	8.24	116.7	174.5	62.0	69.8
Val83	8.26	123.0	176.3	62.4	33.0
Thr84	8.29	118.4	174.6	62.0	69.8
Gln85	8.48	122.9	-	55.7	29.5
Glu87	8.39	121.8	176.7	56.7	30.5
Ala88	8.29	124.6	178.0	52.9	19.2
Gln89	8.24	119.1	176.0	56.0	29.4
Lys90	8.21	122.6	176.1	56.3	33.1
Ala91	8.27	125.4	177.3	52.4	19.5
Glu92	8.33	121.7	174.5	56.7	30.0
Glu94	8.54	120.9	176.5	56.7	30.4
Lys95	8.39	121.9	176.4	56.7	33.0
Ser96	8.36	122.6	-	58.5	63.8
Val99	8.13	120.4	176.3	62.4	32.8
Val100	8.23	124.4	175.7	62.3	32.9
Asp101	8.38	124.5	175.7	54.3	41.3
Ala102	8.20	124.9	177.5	52.5	19.4
Lys103	8.32	120.5	176.4	56.3	32.8
Val104	8.08	121.6	175.8	62.0	32.9
Glu105	8.48	126.6	174.4	54.3	30.0
Gln107	8.52	120.5	176.1	55.9	29.4
Lys108	8.39	122.8	-	56.2	33.1
Gln112	8.30	120.7	175.3	55.5	29.6
Ala113	8.27	126.7	-	50.6	18.2
Glu117	8.50	122.9	176.1	56.3	30.2
Glu118	8.49	123.8	174.5	56.3	30.2
Ser220	8.45	115.6	173.8	58.5	63.9
Ala221	8.07	126.0	174.8	50.5	18.4
Ala122	8.28	123.3	176.3	52.5	19.3

Residue	δ(**^{1}H^{N}**) [ppm]	δ(**^{15}N**) [ppm]	δ(**^{13}C$^{'}$**) [ppm]	δ(**^{13}C$^{\alpha}$**) [ppm]	δ(**^{13}C$^{\beta}$**) [ppm]
Ser123	8.20	114.9	173.8	58.0	63.9
Ala124	8.23	127.0	175.3	50.6	18.4
Ala126	8.40	124.4	177.7	52.4	19.2
Ala127	8.34	123.6	177.9	52.5	19.3
Ser128	8.32	115.0	174.8	58.3	63.8
Ser129	8.36	117.9	174.0	58.2	63.8
Glu130	8.39	123.8	175.6	56.1	30.3
Ala131	8.38	127.0	175.3	50.5	18.2
Lys133	8.52	122.1	177.0	56.2	33.1
Thr134	8.20	115.2	174.2	61.6	69.9
Ser135	8.35	117.9	174.4	58.2	63.8
Glu136	8.39	123.8	-	54.4	30.3
Ser138	8.46	116.1	175.0	58.4	63.8
Ser139	8.44	118.1	174.4	58.6	63.8
Asp140	8.32	122.3	176.2	54.4	41.2
Ala141	8.17	124.6	178.0	52.9	19.2
Lys142	8.23	119.9	176.6	56.4	32.9
Ala143	8.17	124.5	177.7	52.6	19.2
Ser144	8.21	115.1	-	58.3	63.9
Ser147	8.45	116.2	174.7	58.4	63.8
Glu148	8.46	122.6	176.5	56.5	30.1
Ala149	8.29	124.8	177.8	52.7	19.4
Thr150	8.05	113.3	174.0	61.7	69.9
Ala151	8.27	128.0	175.4	50.6	18.3
Ser153	8.44	116.3	174.7	58.4	64.0
Lys154	8.40	123.6	176.3	56.3	33.2
Ala155	8.34	125.2	177.5	52.7	19.4
Asp156	8.27	119.5	176.0	54.3	41.3
Asp157	8.21	121.0	176.7	54.6	41.1
Lys158	8.32	121.8	177.2	56.6	-
Ser159	8.35	116.6	175.1	59.3	63.6
Lys160	8.22	122.8	176.9	56.6	33.0
Glu161	8.31	121.5	176.5	56.8	30.1
Glu162	8.40	121.8	177.2	56.9	30.3
Gly163	8.37	109.5	174.3	45.5	-
Glu164	8.22	120.7	176.4	56.6	30.3
Ala165	8.29	125.0	-	52.7	19.2
Glu169	8.47	123.5	-	56.6	30.2
Ala172	8.43	124.5	177.7	52.4	19.3
Thr173	8.17	115.9	172.9	61.8	69.8
Thr179	8.28	116.4	174.4	62.0	69.8
Lys180	8.42	124.4	176.4	56.3	33.2
Ser181	8.38	117.6	174.5	58.4	63.8
Glu182	8.49	123.2	176.2	56.4	30.0
Val183	8.15	121.2	175.5	62.0	32.8

Residue	δ(**$^{1}H^{N}$**) [ppm]	δ(**^{15}N**) [ppm]	δ(**$^{13}C'$**) [ppm]	δ(**$^{13}C^{\alpha}$**) [ppm]	δ(**$^{13}C^{\beta}$**) [ppm]
Ala184	8.39	129.5	175.3	50.4	18.2
Ala186	8.46	124.6	178.1	52.5	19.3
Ser187	8.35	115.0	174.4	58.4	63.8
Asp188	8.34	122.3	176.2	54.2	41.2
Ser189	8.15	115.8	174.1	58.4	63.8
Lys190	8.29	124.3	174.5	54.3	32.6
Ser192	8.51	116.4	174.9	58.4	63.8
Ser193	8.40	117.8	-	58.5	63.8
Val204	8.16	123.1	175.8	62.3	32.8
Ala205	8.37	127.9	177.5	52.5	19.3
Ala206	8.30	123.6	177.9	52.6	19.3
Thr207	8.04	113.0	174.1	61.6	69.8
Ala208	8.22	126.6	176.9	52.2	18.1
Ala209	8.27	125.0	175.6	50.5	18.2
Thr213	8.18	115.6	174.3	61.9	69.7
Ala214	8.20	126.6	177.6	52.6	19.4
Lys215	8.27	121.0	176.4	56.2	33.2
Ala216	8.36	125.7	177.6	52.6	19.4
Ser220	8.45	115.6	174.1	58.5	63.9
Ala221	8.29	122.8	-	50.5	18.4
Glu225	8.39	122.2	175.9	56.4	30.2
Ala226	8.31	125.8	177.3	52.3	19.4
Lys227	8.32	122.3	174.6	56.6	32.7
Glu230	8.30	122.5	175.8	56.1	30.5
Ala231	8.30	126.7	175.3	50.6	18.1
Ala233	8.46	124.6	178.0	52.5	19.3
Thr234	8.18	113.2	174.3	61.8	69.9
Asn235	8.48	121.0	175.4	53.3	39.0
Ser236	8.36	116.4	-	58.7	63.7
Thr239	8.25	115.7	174.5	62.4	69.7
ILE240	8.06	123.3	175.6	61.0	38.8
Ala241	8.33	128.6	177.3	52.3	19.3
Val242	8.13	120.3	175.9	62.3	32.8
Gln243	8.44	124.9	174.6	55.6	30.0
Asp244	8.04	128.3	177.7	55.9	42.2

The combined use of the $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ secondary chemical shifts for detection of preferences for secondary structure elements (see figure 4.16) indicated two stretches with α-helical preferences starting from residue Lys18 to residue Lys24 (helix α_1) and from residue Asp156 to residue Ser159 (helix α_2), respectively. The predictions obtained by TALOS [25] unambiguously suggests residue Lys18 to residue Lys20 and around Lys158 for α-helical conformation, respectively, but not explicitly the other residual preferences derived from the experimental

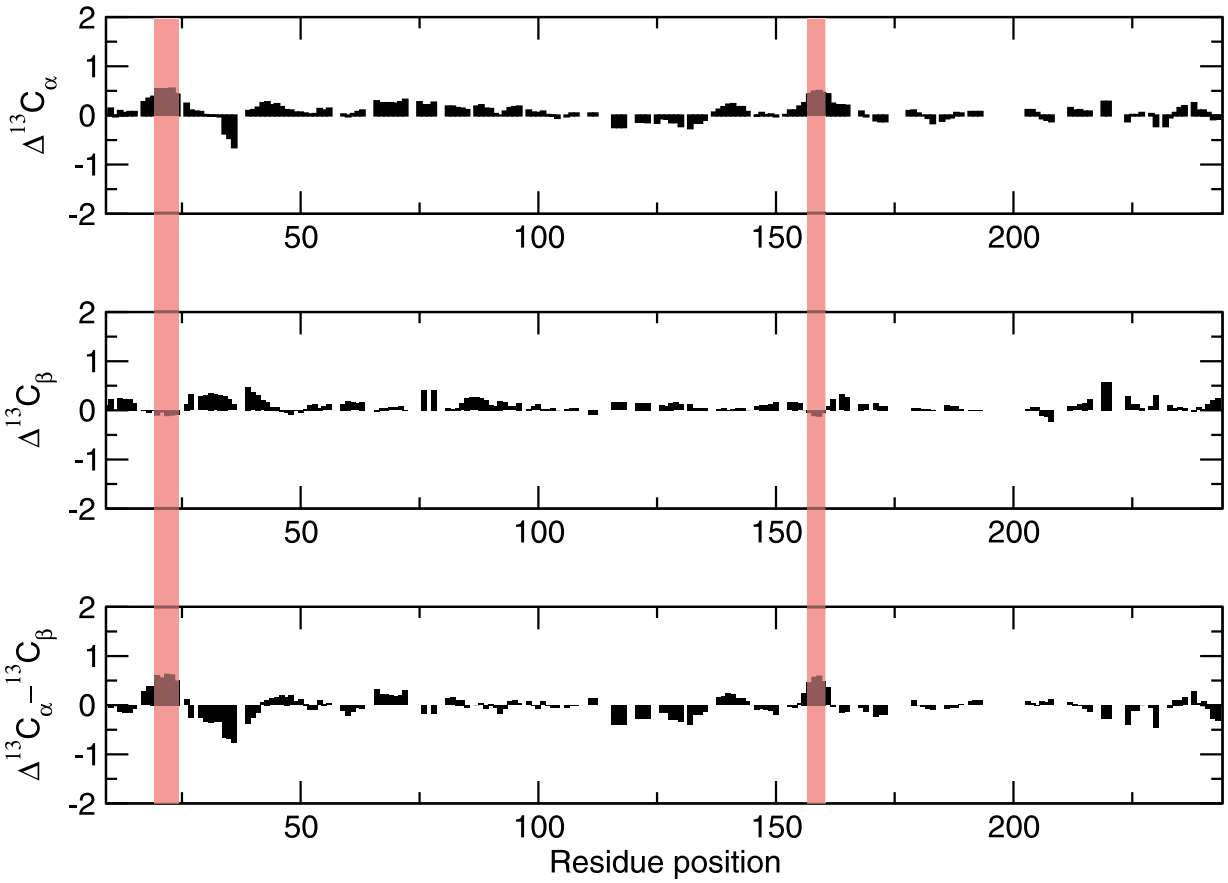


Figure 4.16: $\Delta^{13}C^{\alpha}$, $\Delta^{13}C^{\beta}$, and $\Delta^{13}C^{\alpha}$-$\Delta^{13}C^{\beta}$ secondary chemical shifts as a function of the residue position for BASP1. Bars drawn in red transparent indicate the two regions with α-helical conformational preferences.

data. Indication for an elongated conformation was solely found for residues preceding a proline residue type.

4.2.2 Osteopontin

The $^{1}H^{N}$-^{15}N HSQC of osteopontin (figure 4.17) show all characteristically spectral attributes associated with an intrinsic unfolded protein including residue type peak clustering and a small $^{1}H^{N}$ chemical shift dispersion (0.9 ppm in this case). However, although the molecular size of the analyzed osteopontin construct (219 residues) is roughly the same as for BASP1 (245residues) studied before the extent of completely resolved cross peaks in the $^{1}H^{N}$-^{15}N HSQC spectrum of osteopontin is much higher than for the former case, partly due to a higher number of aromatic residue in the primary sequence (6 phenylalanines and 2 tyrosines for osteopontin compared to only one tyrosine in the entire wild type sequence of BASP1).
Therefore a meaningful site specific ^{15}N relaxation analysis for osteopontin could be performed.

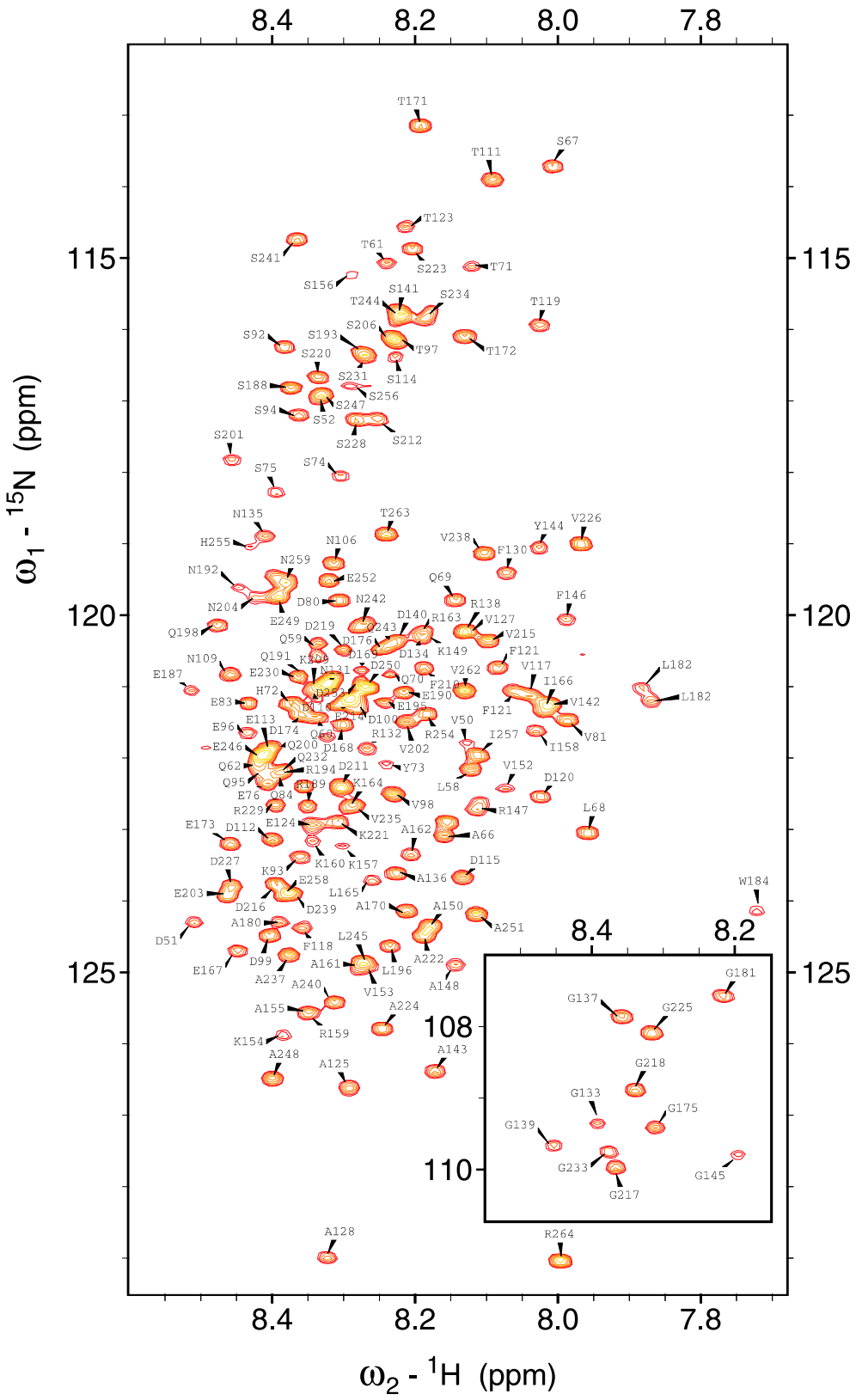


Figure 4.17: 800 MHz 2D $^{1}H^{N}$-^{15}N HSQC of *quail* osteopontin.

The analyzed $^{1}H^{N}$-^{15}N HSQC cross peaks related residues appear well distributed along the whole amino acid sequence, thus almost the entire protein was probed by the dynamical studies derived from T_1, T_2, and NOE measurements.

In total 183 of 207 non-proline $^{1}H^{N}$ and ^{15}N backbone chemical shifts were assigned (88.4%). $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ chemical shifts were allocated for 199 residues (90.9%). Peak positions of assigned carbonyl $^{13}C'$ resonances could be unambiguous identified for 178 out of 219 residues (81.3%). The higher extent (about 20%) of the osteopontin backbone resonance assignment compared to those obtained for the former case BASP1 is explained by the substantial better spectral resolution in the $^{1}H^{N}$ and ^{15}N chemical shift plane.

However, for a N-terminal stretch containing residue His46 to residue His49, no signal resonances were observed, presumably due to fast conformational changes, as indicated by very low signal intensity of residue Val50 in the $^{1}H^{N}$,^{15}N HSQC spectra. Furthermore, only the initial two residue of a poly aspartic stretch starting with residue Asp98 to residue Asp105 could be assigned unambiguously (residues Asp100 to Asp104 seem to be located at a signal patch around 8.28 and 121.2 ppm). The remaining 16 unassigned sequential positions included residues Gln53 to His57, Asn63, Asp64, Glu67, Asp89, Val90, Gln178, Glu190, His199, Arg207, Asn260 and Gln264, which could not be allocated unambiguously because of signal locations in overcrowded regions in the $^{1}H^{N}$ and ^{15}N dimensions.

The assignments have been deposited in the BioMagResBank (http://www.bmrb.wisc.edu) under BMRB accession number 15519 (see also table 4.15).

Table 4.15: Backbone chemical shift assignment of osteopontin (OPN)

residue	δ(**H^{N}**) [ppm]	δ(**^{15}N**) [ppm]	δ(**$^{13}C'$**) [ppm]	δ(**$^{13}C_{\alpha}$**) [ppm]	δ(**$^{13}C_{\beta}$**) [ppm]
HIS49	-	122.6	-	-	-
VAL50	8.13	121.8	175.8	62.4	32.9
ASP51	8.51	124.3	176.5	54.4	41.3
SER52	8.33	116.9	175.2	58.5	63.9
LEU58	8.12	122.3	177.5	55.6	42.2
GLN59	8.34	120.5	176.2	56.2	29.1
GLN60	8.34	121.4	176.4	56.3	29.3
THR61	8.24	115.1	174.8	62.3	70.0
GLN62	8.42	122.1	176.5	55.9	29.3
LEU65	8.12	122.2	178.1	56.1	41.9
ALA66	8.16	123.1	178.7	53.6	18.8
SER67	8.01	113.7	175.1	59.1	63.6
LEU68	7.96	123.1	177.7	55.7	42.3

residue	δ(H^N) [ppm]	δ(^{15}N) [ppm]	δ(^{13}C') [ppm]	δ(^{13}C$_\alpha$) [ppm]	δ(^{13}C$_\beta$) [ppm]
GLN69	8.14	119.8	176.1	56.2	29.3
GLN70	8.24	120.8	176.1	56.0	29.4
THR71	8.12	115.1	174.1	62.1	69.9
HIS72	8.37	121.2	174.2	55.5	29.2
TYR73	8.24	122.1	175.7	58.0	39.1
SER74	8.30	118.0	174.4	58.0	64.1
SER75	8.40	118.3	174.7	58.5	63.8
GLU76	8.41	122.3	176.6	56.5	29.9
ASN78	8.38	119.6	174.7	53.4	39.1
ALA79	8.18	124.3	177.2	52.5	19.4
ASP80	8.30	119.8	175.8	54.4	41.2
VAL81	7.99	121.5	176.8	59.7	32.8
GLU83	8.43	121.2	176.3	56.7	29.8
GLN84	8.39	122.2	-	55.8	29.7
ASP86	8.33	120.4	175.4	54.4	41.4
PHE87	8.08	120.8	-	54.5	39.3
SER92	8.38	116.2	174.8	58.4	63.8
LYS93	8.36	123.4	0.0	56.4	33.1
SER94	8.36	117.2	174.6	58.6	63.8
GLN95	8.41	122.2	175.9	55.9	29.6
GLU96	8.43	121.7	176.5	56.6	29.5
THR97	8.22	116.0	174.4	61.6	69.9
VAL98	8.23	122.5	175.6	62.0	33.1
ASP99	8.40	124.5	175.8	54.3	41.4
ASP100	8.28	121.3	176.1	54.4	41.4
ASP105	0.00	124.1	176.2	54.3	41.1
ASN106	8.31	119.3	175.2	53.4	39.2
ASP107	8.36	121.2	-	54.5	39.3
SER108	8.23	116.1	174.4	58.5	63.7
ASN109	8.46	120.8	175.0	53.4	38.8
ASP110	8.34	121.1	176.3	54.5	41.1
THR111	8.09	113.9	174.3	61.8	70.0
ASP112	8.40	123.2	176.3	54.5	41.2
GLU113	8.41	121.9	176.5	56.6	29.9
SER114	8.23	116.2	-	58.5	63.7
ASP115	8.13	123.7	174.1	54.3	41.2
VAL117	8.04	121.1	175.7	62.2	32.9
PHE118	8.36	124.4	175.7	57.6	39.7
THR119	8.03	115.9	173.5	61.5	69.9
ASP120	8.02	122.6	175.2	54.1	41.3
PHE121	8.07	120.9	173.8	55.6	39.3
THR123	8.21	114.6	174.5	62.0	69.7
GLU124	8.34	122.9	175.6	56.2	30.5
ALA125	8.29	126.6	175.3	50.5	18.3
VAL127	8.13	120.3	175.6	62.0	33.0

residue	δ(H^N) [ppm]	δ(^{15}N) [ppm]	δ(^{13}C') [ppm]	δ(^{13}C$_\alpha$) [ppm]	δ(^{13}C$_\beta$) [ppm]
ALA128	8.32	129.0	-	50.4	18.5
PHE130	8.07	119.4	175.3	57.6	39.5
ASN131	8.32	120.9	174.8	53.0	39.2
ARG132	8.27	121.9	176.8	56.6	30.6
GLY133	8.39	109.3	174.0	45.4	-
ASP134	8.19	120.3	176.3	54.4	41.2
ASN135	8.41	118.9	175.2	53.5	38.9
ALA136	8.23	123.6	178.3	53.0	19.1
GLY137	8.36	107.9	174.4	45.5	-
ARG138	8.13	120.2	177.0	56.4	30.8
GLY139	8.46	109.7	174.0	45.5	-
ASP140	8.22	120.3	176.1	54.4	41.2
SER141	8.22	115.8	174.8	58.6	63.8
VAL142	8.01	121.3	175.9	62.5	32.6
ALA143	8.17	126.4	177.4	52.5	19.2
TYR144	8.03	119.1	176.5	58.3	39.0
GLY145	8.20	109.8	174.0	45.4	-
PHE146	7.99	120.1	175.8	58.2	39.5
ARG147	8.11	122.7	175.7	56.0	30.9
ALA148	8.14	124.9	177.7	52.7	19.2
LYS149	8.19	120.3	176.3	56.2	30.9
ALA150	8.18	124.4	177.5	52.8	19.4
HIS151	8.35	118.5	174.7	55.8	29.0
VAL152	8.07	122.4	175.9	62.5	33.0
VAL153	8.27	124.9	176.0	62.2	32.9
LYS154	8.38	125.9	176.2	56.4	33.1
ALA155	8.35	125.6	177.8	52.7	19.3
SER156	8.29	115.3	174.8	58.6	63.8
LYS157	8.30	123.2	176.6	56.4	33.0
ILE158	8.03	121.6	176.2	61.3	38.9
ARG159	8.35	125.6	176.3	56.2	31.0
LYS160	8.34	123.2	176.4	56.4	33.2
ALA161	8.28	124.9	177.4	52.6	19.2
ALA162	8.21	123.4	177.7	52.6	19.2
ARG163	8.19	120.2	176.2	56.2	31.2
LYS164	8.29	122.7	176.3	56.4	33.1
LEU165	8.26	123.7	177.1	55.4	42.4
ILE166	8.02	121.3	176.2	61.1	38.6
GLU167	8.45	124.7	176.1	56.6	30.3
ASP168	8.30	121.5	176.0	54.4	41.4
ASP169	8.28	120.8	176.1	54.4	41.2
ALA170	8.21	124.2	178.1	52.7	19.4
THR171	8.19	113.2	174.9	62.0	69.9
THR172	8.13	116.1	174.6	61.8	69.9
GLU173	8.46	123.2	176.2	56.6	30.0

residue	δ(H^N) [ppm]	δ(^{15}N) [ppm]	δ(^{13}C') [ppm]	δ(^{13}C$_\alpha$) [ppm]	δ(^{13}C$_\beta$) [ppm]
ASP174	8.36	121.4	176.9	54.5	41.3
GLY175	8.31	109.4	174.1	45.6	-
ASP176	8.25	120.5	176.4	54.4	41.4
SER177	8.22	115.9	-	58.5	63.8
ALA180	8.39	124.3	178.3	52.7	19.2
GLY181	8.22	107.6	173.9	45.3	-
LEU182	7.87	121.2	176.5	55.2	42.5
TRP183	7.88	121.0	174.9	57.0	29.9
TRP184	7.72	124.1	173.4	54.8	29.6
LYS186	8.33	121.7	176.9	56.4	33.2
GLU187	8.51	121.1	176.5	56.9	30.1
SER188	8.37	116.8	174.4	58.6	63.9
ARG189	8.35	122.7	176.2	56.4	30.6
GLU190	8.22	121.1	176.0	56.6	30.2
GLN191	8.35	120.9	175.9	55.8	29.5
ASN192	8.45	119.6	175.3	53.5	39.1
SER193	8.27	116.3	174.8	58.5	63.8
ARG194	8.39	122.2	-	56.0	30.7
GLU195	8.24	121.2	176.1	56.3	30.1
LEU196	8.23	124.6	175.2	53.1	42.2
GLN198	8.48	120.1	-	56.0	-
GLN200	8.39	121.9	175.8	55.8	29.7
SER201	8.46	117.8	174.6	58.3	63.8
VAL202	8.21	121.5	176.2	62.3	32.8
GLU203	8.46	123.9	-	56.7	29.8
ASN204	8.42	119.7	-	53.4	39.0
ASP205	8.31	120.9	176.4	54.5	41.3
SER206	8.23	116.2	174.4	58.6	63.8
LYS209	8.34	121.1	176.4	56.1	33.1
PHE210	8.19	120.7	175.1	57.5	39.8
ASP211	8.30	122.4	176.1	54.0	41.3
SER212	8.26	117.3	-	58.6	63.8
GLU214	8.31	121.2	176.6	56.6	30.1
VAL215	8.10	120.4	176.0	62.3	33.0
ASP216	8.39	123.8	176.7	54.4	41.3
GLY217	8.37	110.0	175.0	45.6	-
GLY218	8.34	108.9	174.1	45.5	-
ASP219	8.30	120.5	176.7	54.4	41.4
SER220	8.34	116.7	175.0	58.6	63.8
LYS221	8.31	122.9	176.7	56.4	33.0
ALA222	8.19	124.5	178.0	52.8	19.4
SER223	8.20	114.9	174.4	58.4	63.8
ALA224	8.25	125.8	178.1	52.8	19.4
GLY225	8.32	108.1	174.3	45.4	-
VAL226	7.97	119.0	176.0	62.4	32.8

residue	δ(H^N) [ppm]	δ(^{15}N) [ppm]	δ(^{13}C') [ppm]	δ(^{13}C$_\alpha$) [ppm]	δ(^{13}C$_\beta$) [ppm]
ASP227	8.46	123.8	176.4	54.3	41.2
SER228	8.28	117.3	175.0	58.6	63.7
ARG229	8.39	122.6	176.8	56.8	30.6
GLU230	8.36	120.9	176.8	56.9	30.1
SER231	8.27	116.4	174.7	58.6	63.8
GLN232	8.39	122.2	176.5	56.0	29.3
GLY233	8.38	109.8	-	45.4	-
SER234	8.19	115.8	174.2	58.4	63.8
VAL235	8.29	122.7	-	61.1	33.0
ALA237	8.38	124.7	177.9	52.5	19.3
VAL238	8.10	119.1	175.8	62.3	33.1
ASP239	8.37	123.9	176.2	54.1	41.3
ALA240	8.31	125.4	178.1	52.9	19.2
SER241	8.37	114.7	174.6	59.1	63.7
ASN242	8.28	120.2	175.2	53.4	38.9
GLN243	8.23	120.4	176.4	56.0	29.6
THR244	8.22	115.8	174.6	61.9	69.9
LEU245	8.27	124.9	177.4	55.3	42.5
GLU246	8.41	122.0	176.5	56.7	29.9
SER247	8.33	116.9	174.4	58.2	63.8
ALA248	8.40	126.5	178.0	52.9	19.3
GLU249	8.39	119.7	176.5	56.9	30.2
ASP250	8.27	121.0	176.2	54.4	41.5
ALA251	8.11	124.2	178.1	52.9	19.3
GLU252	8.32	119.5	176.5	57.0	30.2
ASP253	8.28	120.9	176.6	54.4	41.1
ARG254	8.19	121.4	176.5	56.5	30.5
HIS255	8.43	119.1	175.0	56.1	28.9
SER256	8.28	116.8	174.7	58.6	63.7
ILE257	8.11	121.9	176.4	61.5	38.7
GLU258	8.38	123.9	176.2	56.7	30.1
ASN259	8.38	119.5	-	53.4	38.9
VAL262	8.13	121.1	176.3	62.4	33.0
THR263	8.24	118.9	173.6	62.0	69.9
ARG264	8.02	129.0	-	57.8	30.5

For indication of possible secondary structure element preferences sequence corrected $\Delta\delta^{13}C_\alpha - \Delta\delta^{13}C_\beta$ chemical shift values are plotted against amino acid position (figure 4.18). Non surprisingly, the main part of the backbone resonances are found next to the tabulated random coil values.

For a minor number of residue positions a slight weighting to α-helical conformation was found, as for residue Leu58, residue Leu65 to residue Leu68, residue Arg229, and residue H255. Until

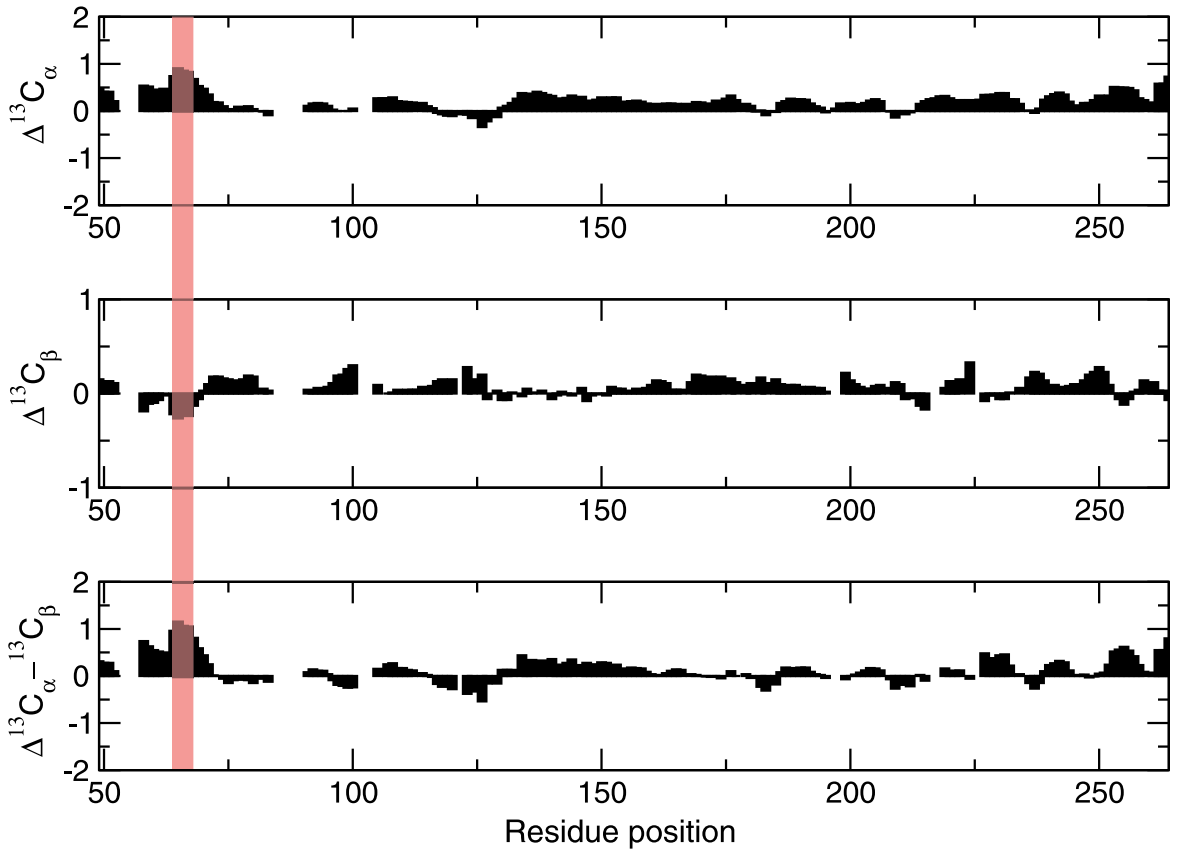


Figure 4.18: $\Delta^{13}C^{\alpha}$, $\Delta^{13}C^{\beta}$, and $\Delta^{13}C^{\alpha}$-$\Delta^{13}C^{\beta}$ secondary chemical shifts as a function of the residue position for osteopontin. A part clearly indicated to show a distinct preference for α-helical conformation is marked in red transparent.

residue H255, these findings were found in close agreement with the (unambiguously defined) predictions inferred from TALOS [25]. Additionally, at this residue positions sequential H^N-H^N NOE contacts could be observed in the 3D ^{15}N-edited NOESY-HSQC experiment. Except of residues preceding a proline residue type not any $\Delta\delta^{13}C_\alpha - \Delta\delta^{13}C_\beta$ chemical shift value below -0.5 was obtained. However, high probability TALOS predictions suggesting a pronounced propensity for the β region of the ϕ,ψ space were found for residues Thr71, Thr97, Thr119, and Ser141 as well as for the fragments residue Val152 to residue Lys154 values and Lys157 to Ile158, respectively.

4.2.2.1 ^{15}N Relaxation

The conformational dynamics of osteopontin was probed by measurement of longitudinal T_1 values, transverse T_2 values, and heteronuclear steady state NOE $^{15}N\{^1H^N\}$ attenuation factors. These parameters are sensitive to dynamics on the picosecond to nanosecond time scale, whereas T_2 can also be effected by conformational or chemical exchange processes on the mil-

lisecond to microsecond time scale as well.

Despite sparse $^{1}H^{N}$ and ^{15}N chemical shift dispersion of osteopontin, in total for 122 out of 183 assigned non proline residues ^{15}N relaxation data could be collected at 298 K and 600 MHz. The T_1 values (see figure 4.19, top) show a relatively strict monotone featureless distribution along amino acid sequence, except for the c-terminal end as a result of increased conformational flexibility, with an average value of 619 ms. T_1 measurements typically found for folded proteins of comparable size are within the range of 840 to 1000ms, which (non surprisingly) indicates an increased overall tumbling of osteopontin relative to its folded counterparts.

In contrast, the measured T_2 values (figure 4.19, middle) show a more sequence depending profile ranging from 114 ms to 747 ms. This finding excludes a complete intrinsic random structure without any sterical restrictions and structural preferences. In such case one would expect a more uniform pattern of the T_2 measurements with increased values at the termini and somewhat more decreased values toward the center of the primary sequence.

In osteopontin (beside residue located in the sequential middle of the protein) several regions can be found which exhibit generally lower T_2 values. First, a fragment next to the N-terminal containing residue Leu58 to residue S75, which enclose the predicted α-helical tendency for residues Leu65 to Leu68. Next, residue Phe118 to residue Phe121 containing Thr119 suggested by TALOS to exhibit a preferency for the β-strand region of the ϕ,ψ conformational space, and residue Thr123 to residue Arg132. Another stretch with decreased T_2 values encompasses residues Leu182, Trp183, and Trp184, whose motion are more sterically constrained because of the spacious indole and extended aliphatic side chains.

These findings are in good agreement with the obtained steady state NOE $^{15}N\{^{1}H^{N}\}$ attenuation factors (figure 4.19.C), that found to be generally higher and strict positive in these ranges. Highest values (0.2 to 0.37) are observable in the middle of the sequence and for Trp184 (0.35). Extensive patches of negative attenuation factors are apparent from residues Ala170 to Gly175, and more pronounced along a 9 residue stretch comprising residues Gly217 to Gly225, a motif build up by residues of small side chains (3 glycines, 2 alanines, 2 serines). In addition, these parts show higher T_2 values about 350 to 450 ms. The total average of T_2 is 280 ms.

4.2.2.2 Spectral Density Mapping

In studies of folded (globular) proteins ^{15}N relaxation data is commonly interpreted by employing the Lipari-Szabo 'model free' approach [103], [104]. However, in cases of unfolded polypep-

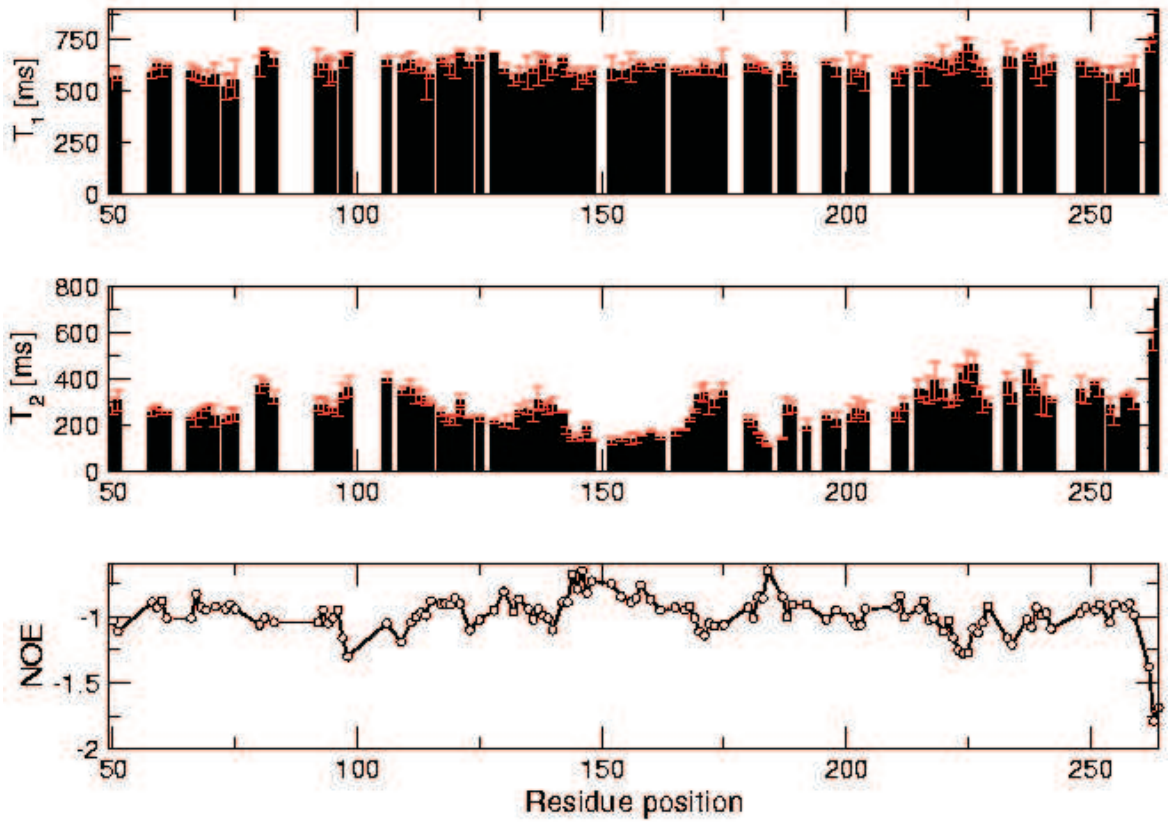


Figure 4.19: Experimental T_1, T_2, and $^{15}N\{^1H^N\}$ NOE measurements of osteopontin at 298K and 600MHz fi eld strength.

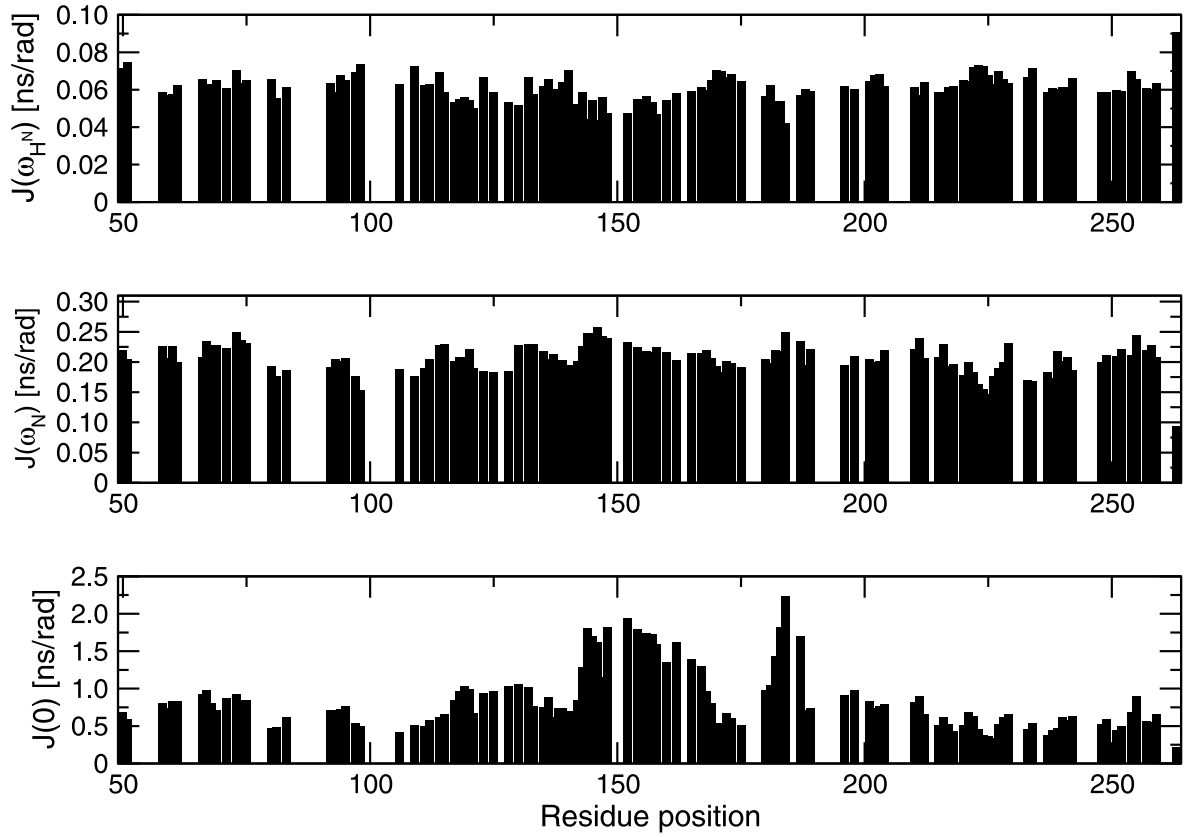


Figure 4.20: Spectral density function of osteopontin at 298K derived at frequencies of 0, 60.8 MHz (ω_N), and 600.1 MHz (ω_{H^N}).

tides this method (which assumes a single isotropic overall correlation time τ_c) is of doubtfully validity because the motions of the backbone ^{15}N$\{^1$H^N vectors are expected to be more complex. Therefore, the relaxation data was analyzed using the spectral density mapping method described in section 2.1.10. The calculated values of $J(\omega)$ are shown in figure 4.20.

The results closely resembles the inverse profiles of the T_1 and T_2 values. For the very most residues $J(0)$ is conspicuously larger than $J(\omega_N)$ indicating a more deciduously shape of the spectral density function and that also low frequency components are contributing to relaxation of osteopontin to some extent. Most restricted motions are observed for amino acids located in the middle of the sequence (residue Ala143 to residue Ala162) and for a patch of residues with spacious aromatic and extended aliphatic side chains encompassing residues Leu182, Trp183, Trp184, Pro185, and Lys186. Another $J(0)$ values distinct larger than average concerning residues Leu65 to residue Leu68, a stretch indicated by secondary chemical shifts to have some propensity for an α-helical conformation (which is additionally found in close conformity with the TALOS analysis [25]), two segments encompassing residue F118 to residue F121 and residue T123 to residue R132 (containing Pro126 and Pro129), both suggested by chem-

ical shift homology search to populate the β-strand region of the ϕ,ψ conformational space. Conversely, parts in the primary sequence were the high frequency component $J(\omega_{H^N})$ is observed to be the major contributor to relaxation pertain residue Ala170 to residue Gly175 and residue Ala222 to residue Gly225, two regions rich on small sidechain residue types.

4.3 Automated chemical shift assignment of methyl groups

The performance of the automated assignment routine described in section 2.5.3 was illustrated by using experimental data sets (unassigned sidechain methyl and assigned backbone chemical shifts together with NOE data extracted from 3D ^{13}C- and 3D ^{15}N-edited NOESY-HSQC) of three proteins, the 168 amino acid truncated form of wild type *dog* ICln ('sample B'), 165 residue *human* Cyclophilin D and a 83 residue polypeptide containing the LIM1 domain of *quail* CRP2.

In a two step mode, the program initially generates a summary of the local assignment possibilities prior global classification (of that results) is performed. For all local assignment proposals and for the best found global assignments (subdivided in non-ambiguous, non-ambiguous single, ambiguous, and potentially wrong assignments) assigned peak lists in nmrview and/or sparky format are generated allowing quick interactive manual inspection of the obtained results in the spectra. For instance, assignment ambiguities could potentially be resolved by taking relative peak intensities into account.

The obtained results and the consistency with the priory performed manual assignment is given in table 4.16.

Interestingly, all proposed assignments marked as non-ambiguous or non-ambiguous single were found to be in complete agreement with the related manual assignments. For four methyl bearing residues of ICln either located at the high dynamical β_6/β_7-loop (CH_3^{β} of Ala104 and Ala117) or in the C-terminal α-helix ($CH_3^{\epsilon 1}$ of Met132, and $CH_3^{\gamma 2}$ of Thr134) not any sidechain methyl assignment possibility was suggested by the program because of intense signal overlap. Subsequent *global* verification of the individually obtained *local* assignments (bottom table 4.16) revealed an incorrect ambiguous assignment. In the case of cyclophilin D four observed methyl chemical shifts entries remained without any assignment proposals ($CH_3^{\delta 1,2}$ of Leu22 located in β-barrel, $CH_3^{\delta 1,2}$ of Leu90 at the β_6/β_7-loop, and $CH_3^{\gamma 2}$ of Val127 in loop between the $3/10$-helix and strand β_8) due to extensive signal overlay.

Table 4.16: Comparison of obtained automated and manual methyl chemical shift assignments.

Results using *local* assignment only				
Protein data	**# observed + manually assigned CH_3**	**# of correctly automatically assigned methyl groups (% correct/% to total)**[a]		
		non-ambiguous	**ambiguous**	**total**
CycD	76	50/50 (100.0/65.8)	15/20 (19.7/39.5)	65/70 (85.5/92.1)
ICln	57	36/36 (100.0/63.1)	16/17 (94.1/29.8)	52/53 (98.1/93.0)
LIM1	24	18/18 (100.0/75.0)	1/4 (25.0/16.7)	19/22 (86.4/91.7)
Results after *global* assignment classification				
Protein data	**# observed + manually assigned CH_3**	**# of correctly automatically assigned methyl groups (% correct/% to total)**[a]		
		non-ambiguous	**ambiguous**	**total**
CycD	76	57/57 (100.0/75.0)	8/8 (100.0/10.5)	70/70 (100.0/92.1)
ICln	57	44/44 (100.0/77.2)	9/9 (100.0/15.7)	53/53 (100.0/93.0)
LIM1	24	22/22 (100.0/91.7)	0/0 (0.0/0.0)	22/22 (100.0/91.7)

[a] '% to total' denotes the percentage of a distinct set of obtained assignments (labeled as non-ambiguous or ambiguous) or the total number of obtained assignment proposals to the total number of observed methyl groups in the spectra. '% correct' indicates the extent of correctness of an assignment set.

For five proposed allocations, all marked as ambiguous by the program, the sequential assignment proposition with the best *local* assignment score was incorrect. Conspicuously, in all that cases the second best assignment proposal could be verified to be coincidental with the manual assignment, the input chemical shift coordinates are located in overcrowded regions of the spectra, and both assignment propositions (the first and the second best) were found in close vicinity at the tertiary NMR structure of Cyclophilin D. However, the complete set of these *local* wrong proposals could be identified as incorrect by the program after *global* classification.

For $CH_3^{\epsilon 1}$ of Met38 and $CH_3^{\gamma 2}$ of Thr47 accommodated at the β_6/β_7-loop and the β_7/β_8-loop of the subdomain II of LIM1/CRP2, respectively, no assignment were proposed by the protocol for the same reason as mentioned in the examples above. The final *global* evaluation of the obtained *local* assignment enabled the correction of three incorrect ambiguous assignments, all mistakenly confounded with methyl groups of close spatial proximity.

The results indicate that principally even in the absence of any tertiary structure for structure based filtering of the possible assignments one can get an adequate percentage (ranging from 63 to 75%) of correct non-ambiguous *local* sequential assignment proposals for methyl groups (located mainly in regular secondary structure). Although via *global* classification the extent of assignments marked as unambiguous by the program could be further increased, obviously one could calculate a primary fold basing on the related NOE contacts of the locally assigned sidechain methyl groups (see figure 4.21), which can be subsequent used for

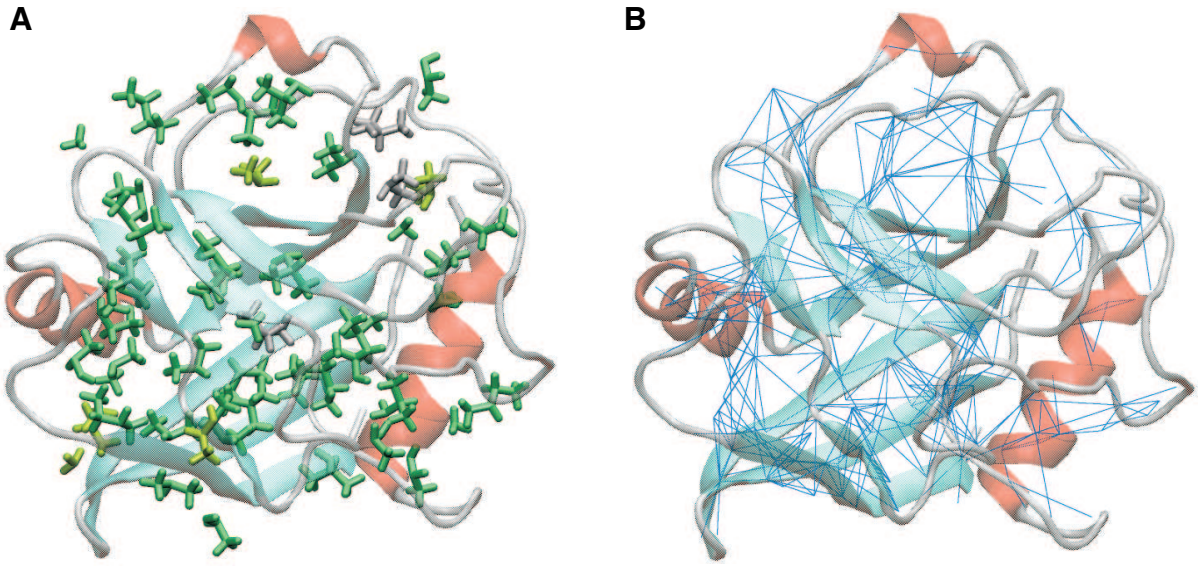


Figure 4.21: (A) Location of the methyl groups visible at the 3D ^{13}C-edited NOESY-HSQC spectrum. Sidechain bonds colored in green refers to methyl groups directly unambiguously assigned by the program using $local$ criteria only, those colored in light green indicate methyl positions additionally assigned by $global$ classifi cation. For methyl groups drawn in gray no assignment proposals were provided. (B) NOE network arising from the assigned set of backbone amides and sidechain methyl groups.

another methyl assignment. However, as mentoined before in the presented examples most of the incorrect ambiguous assignments concerned methyl groups spatial adjacent in the tertiary structure (most likely as consequence of the extensive use of NOE pattern analysis for methyl group assignment constraining), wherefore especially the application of a primary low resolution fold can be expected to be of limitated usability for methyl assignment completion.

4.4 Semi-automated protein structure calculation with a priori sequential backbone resonance assignment

Once both, the sequential assignment of the backbone chemical shift and of sidechain methyl resonances becomes available, their related NOE distances can be used to derive a primarily fold, which can serve as template for a subsequent automated refinement process using well established ARIA routine [102] or similar approaches (see introduction).

The experimental chemical shift and NOESY data of ICln and Cyclophilin D was used to validate the attainable quality of such structure using their NMR model solved manually by conventional analysis as reference. Although both proteins exhibit a comparable β-barrel like fold, especially the sidechain chemical shift dispersion is essentially better for Cyclophilin D, resulting from an higher density of aromatic systems in the structured parts (11 phenylalanines and 3 tyrosines in

Cyclophilin D compared to 4 phenylalanines and 3 tyrosines in the pleckstrin homology domain of ICln). Furthermore, the number of occurring and spectral resolved methyl bearing residues leucine, isoleucine and valine is substantially increased in Cyclophilin D (about 30 compared to ? in ICln) leading to an almost twice extent of NOE distance restraints in the input database. Therefore, in the case of ICln the fold quality approvement was tested when NOE contacts to aromatic sidechain protons are auxiliary implemented for structure calculation. The sequential spin system assignment of these protons could be inferred from 3D ^{13}C edited NOESY-HSQC and (optionally) from (H$^{\beta}$)C$^{\beta}$(C$^{\gamma}$C$^{\delta}$)H$^{\delta}$ and H$^{\beta}$(C$^{\beta}$C$^{\gamma}$C$^{\delta}$)H$^{\delta}$ experiments [198] in a straightforward manner. Moreover, the impact by additional use of backbone H^{N}-Hα NOE distances was examined.

In addition, the quality of a regular β-barrel fold forming Lipocalin Q83 structures were validated generated using idealized (synthetic) NOE input data.

In any cases the (simple) classification of backbone NOEs according their NOESY cross peak volumes was found to be vitally in respect to conformational quality of the obtained structures (data not shown). All results were referenced to those obtained using backbone amide H^{N}-H^{N} NOEs only ('hn'-structures), e.g. without the necessity of (automated) methyl proton sidechain assignment for structure determination, and to the model structures derived by employing the complete set of sidechain NOE contacts. Surface energy graphs of structures generated with and without sidechain NOEs are illustrated in figure 4.23.

A inherent characteristic of backbone NOE distance restraints is their inability in defining the α-helical structure elements to the rest of the molecule. This is indicated by high values in regions of α-helices at the energy fingerprint presentations (figure 4.23). In the final ensemble of ICln derived using backbone H^{N}-H^{N} contacts as the sole source of NOE distance restraints the C-terminal α-helix remain completely disordered. The situation is likewise in the final 'hn' structure bundles of cyclophilin D and lipocalin Q83, although because of insertion of the α-helices between β-strand secondary structure and the occurrence of some more long range NOE contacts within the adjacent connecting loop and linker regions the fuzziness of the α helical parts appear more less pronounced.

Beside a higher resemblance of the β barrel structural parts to those of the reference fine model M_{fine}(derived from a complete set of NOE distance restraints) when additionally applying NOE distances to sidechain methyl groups -CH$_3^{ILV}$, the C-terminal α-helix of ICln (residue Lys126 to residue Leu142) appear proper orientated at the top of the structure by some methyl to backbone amide and sidechain inter methyl NOE contacts at the begin of the α-helical

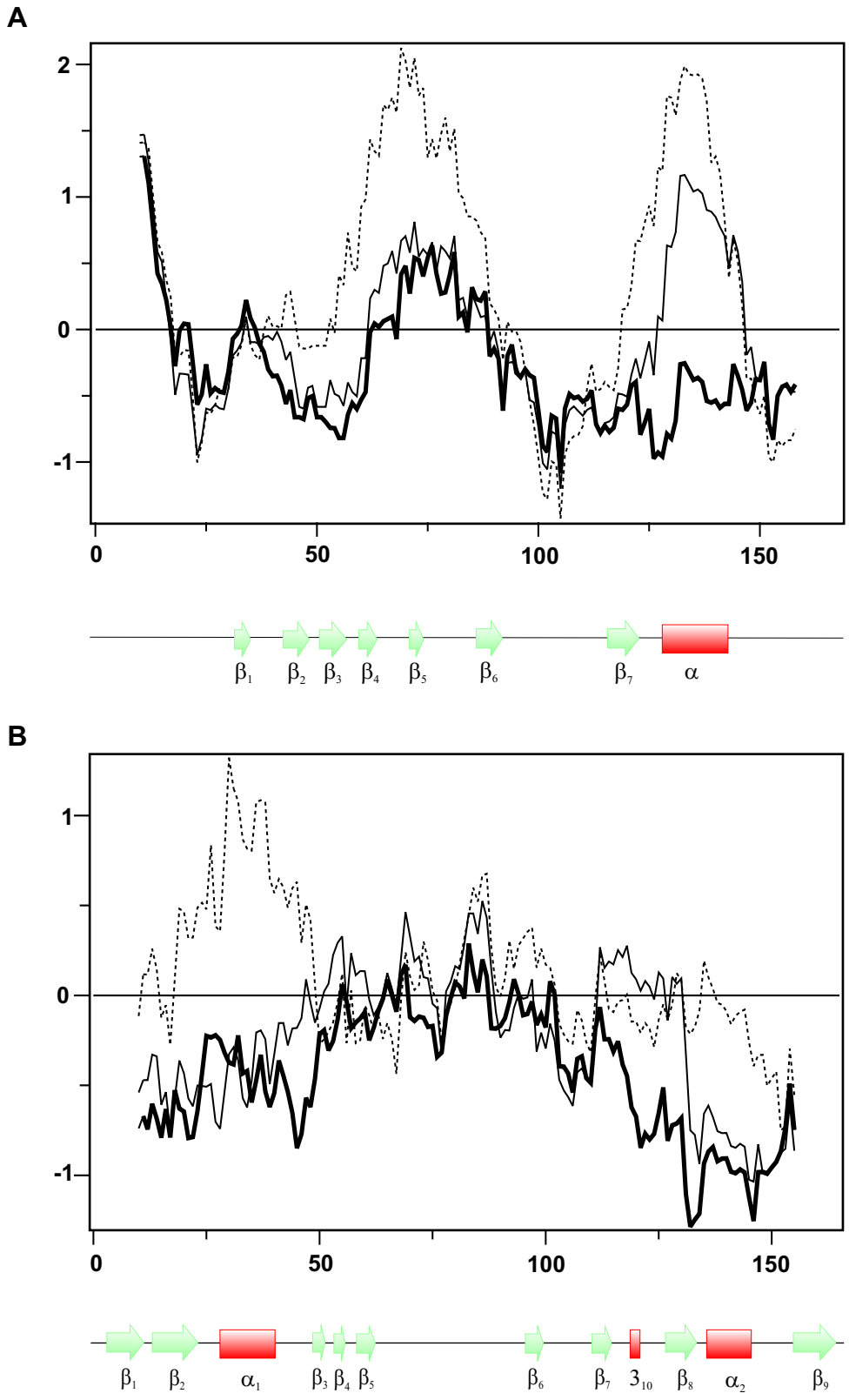


Figure 4.22: Surface energy graphs of the best structure derived using solely backbone H^N-H^N (dashed line), backbone H^N-H^N and sidechain H^N-CH_3^{ILV}/CH_3^{ILV}-CH_3^{ILV} NOE distances (thin line), and a complete set of experimentally NOE distances (bold line) for (A) ICln and (B) cyclophilin D.

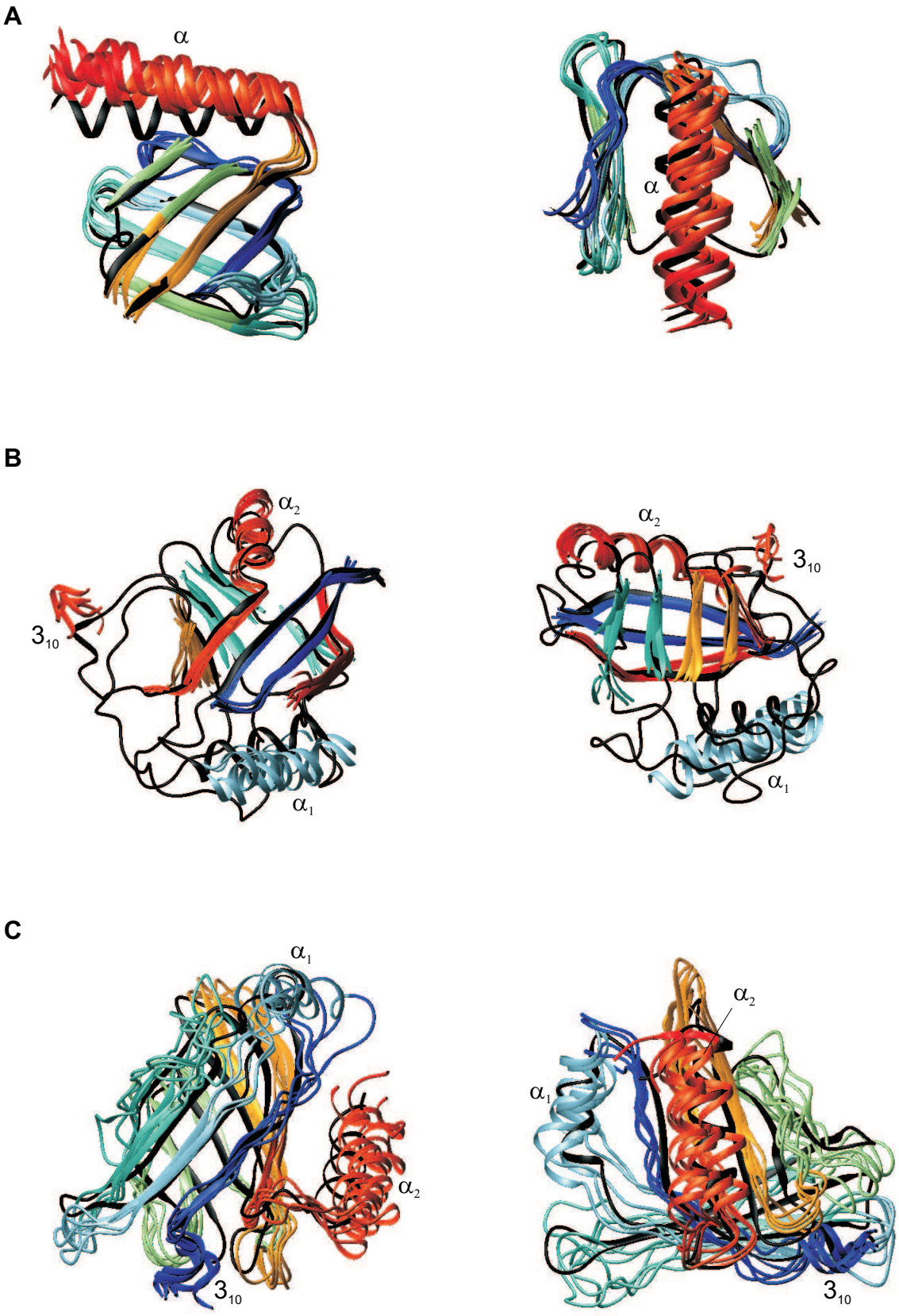


Figure 4.23: Structure ensembles for (A) ICln, (B) Cyclophilin D and (C) lipocalin Q83 generated using backbone H^N-H^N and sidechain H^N-CH_3^{ILV}/CH_3^{ILV}-CH_3^{ILV} NOE distance restraints only (whereas CH_3^{ILV} is related to CH_{3,δ_1}^{Ile}, $CH_{3,\delta_{1,2}}^{Leu}$, and $CH_{3,\gamma_{1,2}}^{Val}$ methyl group pseudo-atoms). A representative set of best five models (individually colored) are superimposed with the reference NMR model (black) using their β-strand atoms for fitting. For clarity, the linker β_4/β_5, loop β_5/β_6, and loop β_5/β_6 of the 'NILV' ICln models, as well as all loop and linker regions of the cyclophilin D 'NILV' structures are not drawn. The

stretch (Leu129($H^{\delta1,2}$)↔Leu87($H^{\delta1,2}$), Leu129($H^{\delta1,2}$)↔Tyr88(H^N), Leu87($H^{\delta1,2}$)↔Phe133(H^N) and Leu87($H^{\delta1,2}$)↔Met132(H^N)) although not so compact defined as when applying a complete set of sidechain NOE contacts due to the absence of methyl NOE distances at the end of the helical part. In Cyclophilin D, the extra use of 12 backbone amide to sidechain methyl NOEs in helix α_2 (residue Met136 to residue Ser141) for structure generation give rise for its adequate positioning to the β-barrel when comparing with the related M_{fine} structure (where 38 long-range NOEs for helix α_2 could be used), whereas the only four helix α_2 H^N/CH_3^{ILV} NOE contacts (Ile78($H^{\delta1,2}$)/Ala38(H^N) and Ile78($H^{\delta1,2}$)/Leu39($H^{\delta1}$)) to the core results in a rather small precision compared M_{fine}, where in total 44 experimentally derived long-range NOE distances were applied for defining that α-helical stretch. The extent of fold quality increasement by the availability of CH_3^{ILV} NOE distances for structure determination in terms of PROSA z- and WHATIF packing scores as well as the conformability of the 'HN' and 'HNILV' structure to their related fine models in terms of RMSD values are given in table 4.17. For ICln the influence of additionally use of backbone amide NOEs to aromatic $H^{\delta1,2}$, $H^{\epsilon1,2}$, and H^{ζ}, and to H^{α}protons, respectively, are shown.

Table 4.17: Quality of obtained structures using sparse input data

Protein # residues	**Selected protons in NOEs distances** [a]	**What IF packing score**	**PROSA z-score** [b]	**RMSD to reference**
ICln (168aa)	H^N	-2.12	-0.30	3.02
	H^N, $CH_3^{(ILV)}$	-1.72	-2.04	1.38
	H^N, $CH_3^{(ILV)}$, $H^{\delta1,2(F)}$/$H^{\epsilon1,2(F)}$/$H^{\zeta(F)}$	-1.55	-2.07	1.24
	H^N, $CH_3^{(ILV)}$, $H^{\delta1,2(F)}$/$H^{\epsilon1,2(F)}$/$H^{\zeta(F)}$, H^{α}	-1.45	-2.38	1.22
	all	-1.14	-3.25	-
Cyclophilin D (165aa)	H^N	-2.11	-3.15	1.48
	H^N, CH_3^{ILV}	-2.13	-5.08	1.29
	all	-1.71	-6.02	-
Lipocalin Q83 (157aa)	H^N	-2.01	-3.54	1.861
	H^N, CH_3^{ILV}	-2.14	-5.13	-
	all	-1.47	-6.84	-

[a] in NOE distance restraints; CH_3^{ILV} emcompasses $CH_3^{\delta1}$ and $CH_3^{\delta2}$ of residue type leucine, $CH_3^{\gamma1}$ and $CH_3^{\gamma2}$ of residue type valine, and $CH_3^{\delta1}$ of isoleucine. [b] in respect to surface energy.

For cyclophilin D and ICln, the application of sidechain CH_3^{ILV} NOEs in addition to backbone H^N-H^N NOE distances improves the structural concordance of their β-barrels to the reference NMR fine model from RMSD 1.48 to 1.29, and 3.02 to 1.24, respectively.

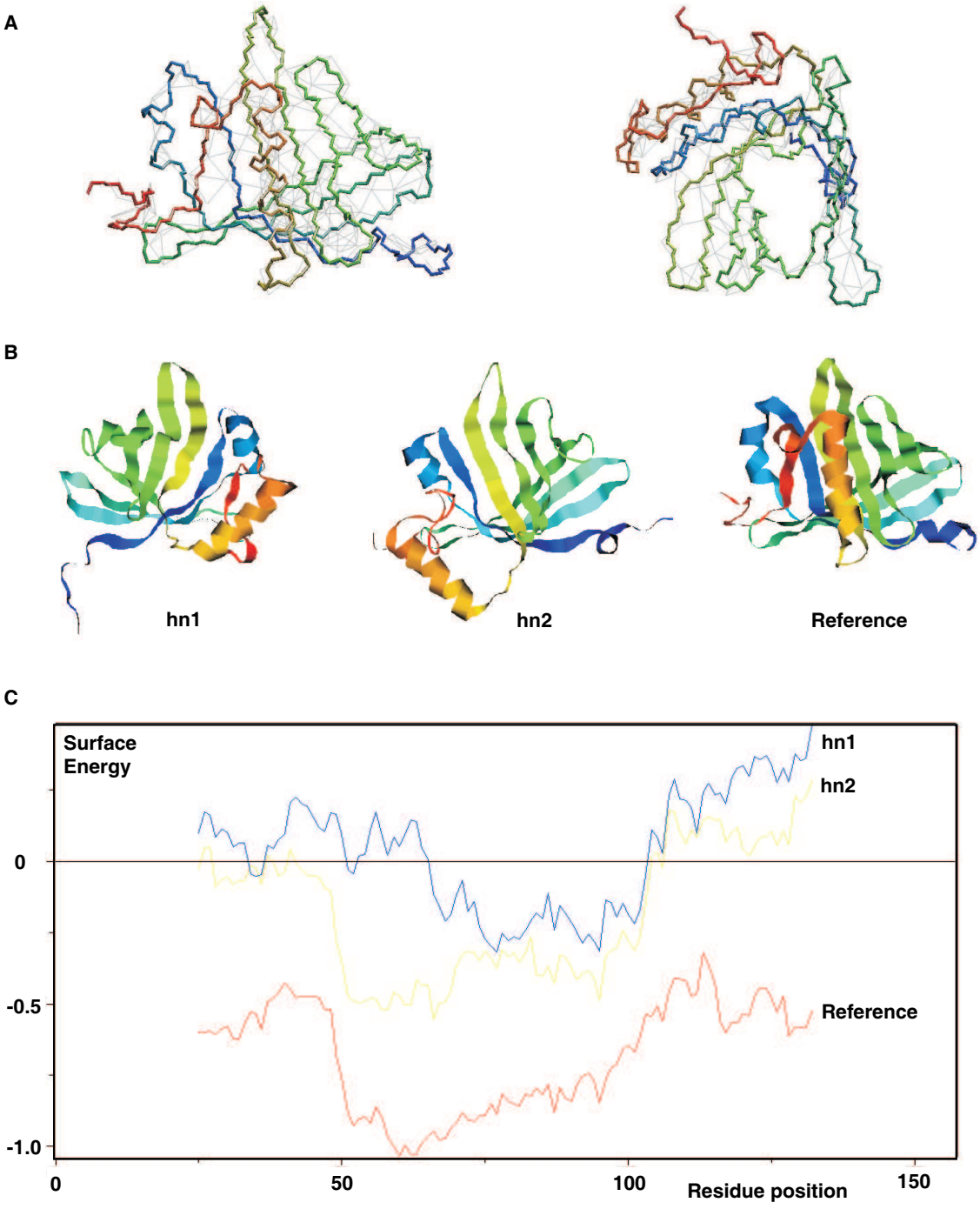


Figure 4.24: (A) Used (backbone) NOE distances (shown in gray) for structure calculation of Q83. (B) Structural comparison of two models (denoted as 'hn1' and 'hn2' offering roughly the same lowest energy) with the published reference NMR structure (1JZU.pdb, termed as 'fine', [59]). (C) Overall surface energy plots of these models applying a window size of 50 residues.

In contrast, for the regular β-barrel structure of lipocalin Q83 no refinement was observed due to ideal synthetic NOE distance input data set, the more regular form of the β-barrel and the associated higher number of inter-strand backbone H^N-H^N NOE contacts.

However, demonstrated at the example of lipocalin Q83 there is another reason for the necessity of sidechain NOEs application for primarily fold generation. Using backbone NOE contacts only, the restraining information is restricted *along* structure elements as α-helices, $beta$-sheets, and $beta$-barrels, but there are no restraints acting towards the core. For lipocalin Q83 even in the presence of chiral restraints this results in the occurrence of 'inside-out' folds with identical XPLOR total energy U_{total} when excluding time consuming electrostatics for explicit solvent. Calculation of the PROSA overall surface energy led to +2.28 kcal/mol for the 'inside-out'-model whereas a value of 1.36 kcal/mol was found for the correct folded model (see also energetic fingerprint in figure), which suggests that the incorrect fold should be distinguishable as long as a high number of correct initial NOE are available, e.g. no other structural inadequateness or errors are present[1]. No certifiable impact on the structural results or the convergence behavior of the md SA calculations of lipocalin Q83 was observed when omitting the single disulphide bridge (between residue Cys59 and residue Cys152).

The presented results show that the application of unambiguous sidechain CH_3^{ILV} NOEs (as feasible derived e.g. by dint of selective labeling techniques) for the semi-automated generation of an adequate initial model for subsequent structure refinement is vital. The data indicate that in principle CH_3^{ILV} functions can provide the NOE contacts requisite for obtaining a model that closely resembles the final structure derived exclusively manually from conventional structure determination.

4.5 Rapid protein fold validation using a direct method in context with residue type specifi c isotope labeling

4.5.1 Extraction of articial intramolecular distance restraints

Starting from initial atomic distribution generated using experimental or structure-based NOE distances only (Simulated Annealing I, round 0), artificial distance restraints A_{ij} had to be extracted in order to achieve a converged final structure of atomic densities (or 'clouds' according [50]). Successive adding of A_{ij} in subsequent round of structure calculation to the input

[1]Note that the PROSA mean force potential is carefully derived from atomic distances found in high quality structures. A classifi cation between a bad and an even more bad structure model is commonly not feasible

database decreases the degree of freedoms of the atomic system and give raise to small σ_{ij} values of distances d_{ij} which are not necessarily adequate defined by the initial input NOEs, thus potentially not compatible with the correct structure.

Low number of experimentally derived restraints in several structural parts (as inherent characteristic in loops or linker regions) and the complete absence of a covalent template systematically promote in overly extended structures and derived distances thereof appear systematically overestimated. Therefore it was found that small standard deviations σ_{ij} of distances d_{ij} alone are not a sufficient criterion. For that reasons more stringent criteria taking input NOE density around the concerned atoms into account have to be applied. The application of the pseudo NOE term $U_{NOEPseudo}$ for weak bias backbone-backbone H^N-H^N NOE distances and NOE contacts of sidechain towards statistically more relevant distance values (about 5-6Å, while practically allowing distances up to 7-8Å accounting for possible effects of spin-diffusion) additionally counteract the tendency of distance overestimation in subsequent rounds and results in significant more conserved distances for subsequent analysis.

The (ltering) effect of the selection criteria for the (experimental) test case cyclophilin D is shown in figure 2.12.

For that example the selected distance set A_{ij}^0 from the initial calculation round 0 contain distances up to about 35Å due to the maximal length l_{max} of 7 individual experimental NOE contacts that build up a NOE path. In subsequent rounds distances connected by a maximal NOE path length l_{max} of 4 were analyzed (corresponding distances up to about 25Å (compare see figure 2.12.A,C) for selection in order to impede generation of artificial restraints hardly non-compatible with the correct native fold, and because evaluation of distances related to longer path length were found to be not demandable to achieve convergence of the procedure.

As consequence of the mode of action of the selection criteria reflects the original density of experimental NOE the number of newly obtained artificial distance restraints A_{ij}^1 in a part of the protein. Hence, the dense network of backbone H^N-H^N NOEs inherent to β barrels or β-sheets give rise for a large number of deduced artificial distance restraints A_{ij} and thereof a high precision for that structural parts in the final structure ensemble. Since typically there are more experimental long range NOEs found in β structure elements than for α-helical regions the backbone chain of α-helices remain generically less defined compared to those of β-strands.

Table 4.18 gives a statistics of the available number of the experimentally derived or structure based generated input NOE used for evaluation of the procedure and the extent of deduced artificial distance restraints A_{ij} for the whole test set of proteins. The number of A_{ij} varies with

the original number of NOE distance restraints and to some extend with the fold topology of the proteins studied. The set of relevant atoms utilized for calculation of the 'cloud' densities are given in the headers. As measure for the restricting properties of a distance restraint series in the individual rounds of structure calculation a rp values is defined as:

$$RP_{mean} = (\sum_{i=1}^{N} lb_i/ub_i) \cdot 100\%/N \cdot \tag{4.1}$$

whereas lb_i and ub_i denotes the lower and upper bound of a restraint i, and N is the total number of restraints. As consequence of successive decreasing σ_{ij} values in subsequent rounds of atomic density model calculation and A_{ij} extraction this values increases.

Table 4.18: Proteins tested and extent of input and deduced artifi cial restraints

Calmodulin backbone (BB) H^N and sidechain (SC) Methyl(I,L,V)				
Restraint classification	*# theoretical NOE's* r_{ij}	*# artificial* d_{ij} *extracted from Round 0*	*Round 1*	$r_{ij} + \sum d_{ij}$ used in *Round 2*
BB-BB	417	202	1087	1706
BB-SC	235	111	817	1163
SC-SC	64	10	127	201
Total	716	323	2031	3070
RP_{mean} [%]	28.6	41.4	53.7	46.6
Violations $>$ 2Å [%]	0	3.4	0.5	0.7
Carbonic anhydrase I backbone (BB) H^N and sidechain (SC) Methyl(I,L,V)				
Restraint classification	*# theoretical NOE's* r_{ij}	*# artificial* d_{ij} *extracted from Round 0*	*Round 1*	$r_{ij} + \sum d_{ij}$ used in *Round 2*
BB-BB	687	5398	5089	11174
BB-SC	502	5131	5053	10686
SC-SC	193	1224	1113	2530
Total	1382	11753	11255	24390
RP_{mean} [%]	26.4	44.4	75.9	57.9
Violations $>$ 2Å [%]	0	1.7	6.6	3.9

Cyclophilin D
exp. backbone (BB) H^N and sidechain (SC) CH_3(I,L,V)

Restraint classification	*# experimental NOE's* r_{ij}	*# artificial* d_{ij} *extracted from*		$r_{ij} + \sum d_{ij}$ used in
		Round 0	*Round 1*	*Round 2*
BB-BB	247	167	549	217
BB-SC	165	96	553	814
SC-SC	50	18	149	963
Total	462	281	1251	1994
RP_{mean} [%]	25.9	43.7	53.4	45.7
Violations $>$ 2Å [%]	0	0.7	4.7	

Cyclophilin D
backbone (BB) H^N and sidechain (SC) CH_3(I,L,V)/$H_{aromatic}$(F,Y)/NH_2(Q,N)

Restraint classification	*# experimental NOE's* r_{ij}	*# artificial* d_{ij} *extracted from*		$r_{ij} + \sum d_{ij}$ used in
		Round 0	*Round 1*	*Round 2*
BB-BB	247	787	1049	2087
BB-SC	285	699	1632	2615
SC-SC	102	200	500	838
Total	638	1686	3181	5540
RP_{mean} [%]	28.1	43.6	56.6	49.4
Violations $>$ 2Å [%]	0	2.6	5.2	4.0

Cyclophilin D
backbone (BB) H^N,H_α and sidechain (SC) CH_3(I,L,V)/$H_{aromatic}$(F,Y)/NH_2(Q,N)

Restraint classification	*# experimental NOE's* r_{ij}	*# artificial* d_{ij} *extracted from*		$r_{ij} + \sum d_{ij}$ used in
		Round 0	*Round 1*	*Round 2*
BB-BB	846	3354	3275	7475
BB-SC	361	1981	2228	4570
SC-SC	218	1066	2474	3758
Total	1425	6401	7977	15803
RP_{mean} [%]	27.4	44.7	62.1	51.9
Violations $>$ 2Å [%]	0	-	-	-

ICln
backbone (BB) H^N,H_α and sidechain (SC) CH_3(I,L,V)/$H_{aromatic}$(F,Y)/NH_2(Q,N)

Restraint classification	*# experimental NOE's* r_{ij}	*# artificial* d_{ij} *extracted from*		$r_{ij} + \sum d_{ij}$ used in
		Round 0	*Round 1*	*Round 2*
BB-BB	336	1612	1219	3157
BB-SC	235	1248	1345	2965
SC-SC	82	747	1041	1891
Total	653	3607	3605	8013
RP_{mean} [%]	31.7	44.4	61.5	50.9
Violations $>$ 2Å [%]	0	0.2	-	-

LIM1(CRP2)
backbone (BB) H^N,H_α and sidechain (SC) CH_3(I,L,V)/$H_{aromatic}$(F,Y)/NH_2(Q,N)

Restraint classification	*# experimental NOE's* r_{ij}	*# extracted artificial* d_{ij} *from Round 0*	*Round 1*	$r_{ij} + \sum d_{ij}$ used in *Round 2*
BB-BB	245	164	2989	3398
BB-SC	127	96	910	1133
SC-SC	92	92	1042	1236
Total	464	352	4941	5767
RP_{mean} [%]	28.4	43.2	53.1	50.6
Violations $>$ 2Å [%]	0	-	-	-

Plastocyanin
backbone (BB) H^N and sidechain (SC) CH_3(I,L,V)

Restraint classification	*# theoretical NOE's* r_{ij}	*# artificial* d_{ij} *extracted from Round 0*	*Round 1*	$r_{ij} + \sum d_{ij}$ used in *Round 2*
BB-BB	262	448	1179	1889
BB-SC	247	392	1184	1823
SC-SC	84	83	297	464
Total	593	923	2660	4176
RP_{mean} [%]	25.6	42.7	52.4	46.5
Violations $>$ 2Å [%]	0	2.1	0.6	0.8

Thioredoxin
backbone (BB) H^N and sidechain (SC) CH_3(I,L,V)

Restraint classification	*# experimental NOE's* r_{ij}	*# artificial* d_{ij} *extracted from Round 0*	*Round 1*	$r_{ij} + \sum d_{ij}$ used in *Round 2*
BB-BB	266	398	1117	1781
BB-SC	206	428	1082	1716
SC-SC	63	104	254	421
Total	535	930	2453	3918
RP_{mean} [%]	24.5	44.9	55.1	48.5
Violations $>$ 2Å [%]	0	1.4	1.3	1.1

4.5.2 Practical considerations and structural results

Applying the structure calculation protocol outlined in section 2.7 convergent atom densities were obtained, which correctly specify the secondary structure elements and their orientation to each other. The quality of these protein images (table 4.19, figure ?) strongly correlate with the extent of the provided input data (number of atoms applied and NOE contacts thereof) and to some degree with the fold topology for reasons mentioned before. Superponable atoms in a final converged cloud ensemble were identied by the analysis of their distance matrix $\mathbf{D}_{ij}$. The occurrence of a high number of conserved distances d_{ij} indicate the presence of structurally dened regions.

Table 4.19: Quality of the obtained unconnected atomic 'cloud' protein images.

Protein	**cloud type** [a]	**RMSD** [b] **[Å]** clouds[1]/to NMR-model[2]	**selected bb H^N atoms**
Calmodulin	set A theo.	2.84 / 2.86	N-term domain
		1.29 / 1.40	C-term domain
		0.73 / 0.75	β-sheet
Carbonic Anhydrase I	set A theo.	1.04 / 1.22	all converged
		0.34 / 0.45	α-helices,β-sheets
		0.67 / 0.74	α-helices
		0.17 / 0.30	α_1,α_3-helices only
		0.12 / 0.26	β-sheets
Cyclophilin D	set A exp.	5.61 / 5.84	all converged
		0.17 / 0.32	α_1-helix
		1.78 / 1.81	α_2-helix
		1.74 / 1.92	β-barrel
	set B exp.	2.62 / 2.76	all converged
		1.11 / 1.31	α-helices,β-sheets
		0.61 / 0.74	α_2-helix only, β-sheets
		0.57 / 0.70	β-barrel
	set C exp.	1.61 / 1.71	all converged
		0.50 / 0.61	α-helices, β-sheets
		0.60 / 0.73	α-helices
		0.41 / 0.50	β-barrel
ICln	set C exp.	0.84 / 1.06	all converged
		0.14 / 0.30	α-helix, β-sheets
		0.08 / 0.28	β-barrel
		0.17 / 0.30	C-term. α-helix
Lim1	set C exp.	1.89 / 2.08	all converged
		1.86 / 1.99	N-term. domain
		1.72 / 1.87	C-term. domain
		1.24 / 1.42	α-helix, β-sheets
		1.16 / 1.31	β-sheets

Protein	**cloud type** [a]	**RMSD** [b] **[Å]** clouds[1]/to NMR-model[2]	**selected bb H^N atoms**
			β-sheets
		1.47 / 1.53	N-term. domain β-sheets
		0.44 / 0.53	C-term. domain
			C-term. β-sheets
		0.95 / 1.09	+ α-helix
Plastocyanin	set A theo.	1.12 / 1.17	all converged
		0.80 / 0.84	β-sheets
Thioredoxin	set A theo.	1.61 / 1.66	all converged
		1.34 / 1.38	α-helix, β-sheet
		1.72 / 1.76	α-helix
		0.26 / 0.29	β-sheet

[a] 'set A theo/exp': atomic cloud comprising H^N, and -CH_3(I,L,V,A,T) atoms; derived by using either their individual theoretical (structure based) or experimental NOEs, respectively.
'set B exp': as 'set A', but supplementary containing aromatic -$H_{aromatic}$(F,Y), and sidechain amide -NH_2(Q,N) atoms; calculated by using experimental NOEs, only.
'set C exp': as 'set B exp', but also including H_α atoms and additionally employing their experimental H^N-H_α NOEs for atomic cloud calculation.
[b] Residue numbers used for superimposing backbone amid H^N atoms. (a) Total Backbone H^N atoms found to be converged. (b) as (a), but located in regular secondary structure only.
Calmodulin:
(a) set A exp.: 1-38, 40-42, 44-65, 67-118, and 120-148.
(b) N-terminal domain (1-56): 6-19(α_1), 26-27(β_1), 29-36(α_2), 45-55(α_3), 63-64(β_2). C-terminal domain (95-148): 65-92(α_4), 102-112(α_5), 118-126(α_6), and 138-144(α_7).
Carbonic Anhydrase I
(a) set A exp.: 5-12, 14-20, 22-29, 31-45, 47-52, 54-82, 84-154, 156-174, 176-180, 182-185, 187-194, 196-200, 203-239, 241-246, 248-249, and 251-260.
(b) 32-33(β_1), 39-40(β_2), 45-50(β_3), 56-61(β_4), 66-70(β_5), 78-82(β_6), 88-97(β_7), 108-109(β_8), 112-113(β_9), 116-124(β_{10}), 131-134(α_1), 141-150(β_{11}), 158-163(α_2), 173-175(β_{12}), 191-196(β_{13}), 207-212(β_{14}), 216-218(β_{15}), 220-227(α_3), and 257-258(β_{16}).
Cyclophilin D
(a) set A exp.: 2-3, 5-15, 17-29, 31-57, 59-64, 66-69, 71-83, 85-94, 96-100, 106-147, 150-165. set B exp.: 2-3, 5-15, 17-29, 31-57, 59-64, 66-69, 71-83, 85-94, 96-100, 103-104, 106-147, 150-165. set C exp.: as set B exp.
(b) 5-12(β_1), 15-24(β_2), 31-39(α_1), 51-52(β_3), 55-57(β_4), 61-64(β_5), 97-100(β_6), 112-115(β_7), 127-133(β_8), 136-145(α_2), and 156-164(β_9).
ICln:
(a) set C exp.: 12-16, 22-29, 31-67, 69-82, 84-96, 116-122, and 124-143.
(b) 33-36(β_1), 44-49(β_2), 53-56(β_3), 62-64(β_4), 72-75(β_5), 86-92(β_6), 116-122(β_7), and C-terminal 128-140(α_1).
Plastocyanin:
(a) set A exp.:1-15, 17-21, 23-35, 37-46, 48-57, 59-65, 67-85, and 87-99.
(b) 1-5(β_1), 14-15(β_2), 18-22(β_3), 25-31(β_4), 40-41(β_5), 69-74(β_6), 78-83(β_7), and 93-99(β_8).
Thioredoxin:
(a) set A exp.:1-33, 35-39, 41-74, and 76-105.
(b) 3-4(β_1), 8-17(α_1), 23-28(β_2), 33-36(α_2), 39-42(α_3), 45-48(α_4), 53-58(β_3), 63-69(α_5), 76-81(β_4), 84-90(β_5), and 95-104(α_6).

Because only a distance related potential U_{NOE} was acting as the sole attractive term in the

applied md simulated annealing protocol the derived atom densities split evenly in their *global* mirror image-like subsets. Due to the absence of a covalent protein template and improper energy terms, which define the stereochemistry around chiral centers, the correct protein fold could not be distinguished from the inverse wrong one (see figure 4.25).

In addition, a α-helix build up by unconnected atoms can potentially adopt both, a right and left handed helical structure resulting in up to 2^n furthermore *local* family members or subsets with degenerate energy. Whereas the local handedness of α-helices could be resolved under favorable circumstances (e.g in α-helices of high spatial resolution), the correct global chirality cannot be a priory distinguished at this stage due to the absence of any force eld term other than distance restraints.

Assuming a large set of standard triple resonance spectra is additionally available (beside NOESY data) together with a plenteous amount of H^{α}/H^{β} NOEs one could try to build chiral centers about the backbone C^{α} atom. Although, provided a broad dispersion of $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and ^{13}C' chemical shifts, in principle the generation of unambiguous intra-residue C'-$C^{\alpha}/H^{\alpha}(C^{\beta}/H^{\beta})$-$N^{H}$ connections should be feasible for a significant number of residues (preferably distributed along the whole primary sequence), in practical defining of a high number of unambiguously experimentally derived initial H^{α}/H^{β} NOE contacts requisite to obtain a sufficient structural resolution and orientation of these fragments to the residual atoms (for identifying the correct global stereoisomere) is quite unrealistic. On the other hand, if residual dipolar couplings are available one could expect a more ameliorated structure using such templates.

Nevertheless, the non-covalent three dimensional proton densities obtained at this stage are of great value. Sometimes one is interested to verify the quality of a homology model, which often suffers due to some uncertainty of the precise location of its secondary structure elements and their exact orientation to each other. The correct fold of the two resulting mirror image-related global atomic 'cloud' sets then could be simple identified using the coordinates of that model, whose tertiary fold is examined by the atomic 'cloud' structure. Furthermore, the atomic densities can be used to support the generation such homology model when applied in combination with database mining and molecular fragment replacement (see figure 4.26).

Another implementations of atomic density 'clouds' is the aid of the sequential signal assignment process and in the validation of ambiguous NOEs using structure based criteria.

The distances extracted from the best structure are found to be very useful for guiding the sequential backbone resonance assignment, if only limited data set of backbone chemical shifts are available. This was demonstrated applying the automated simulated annealing Monte Carlo

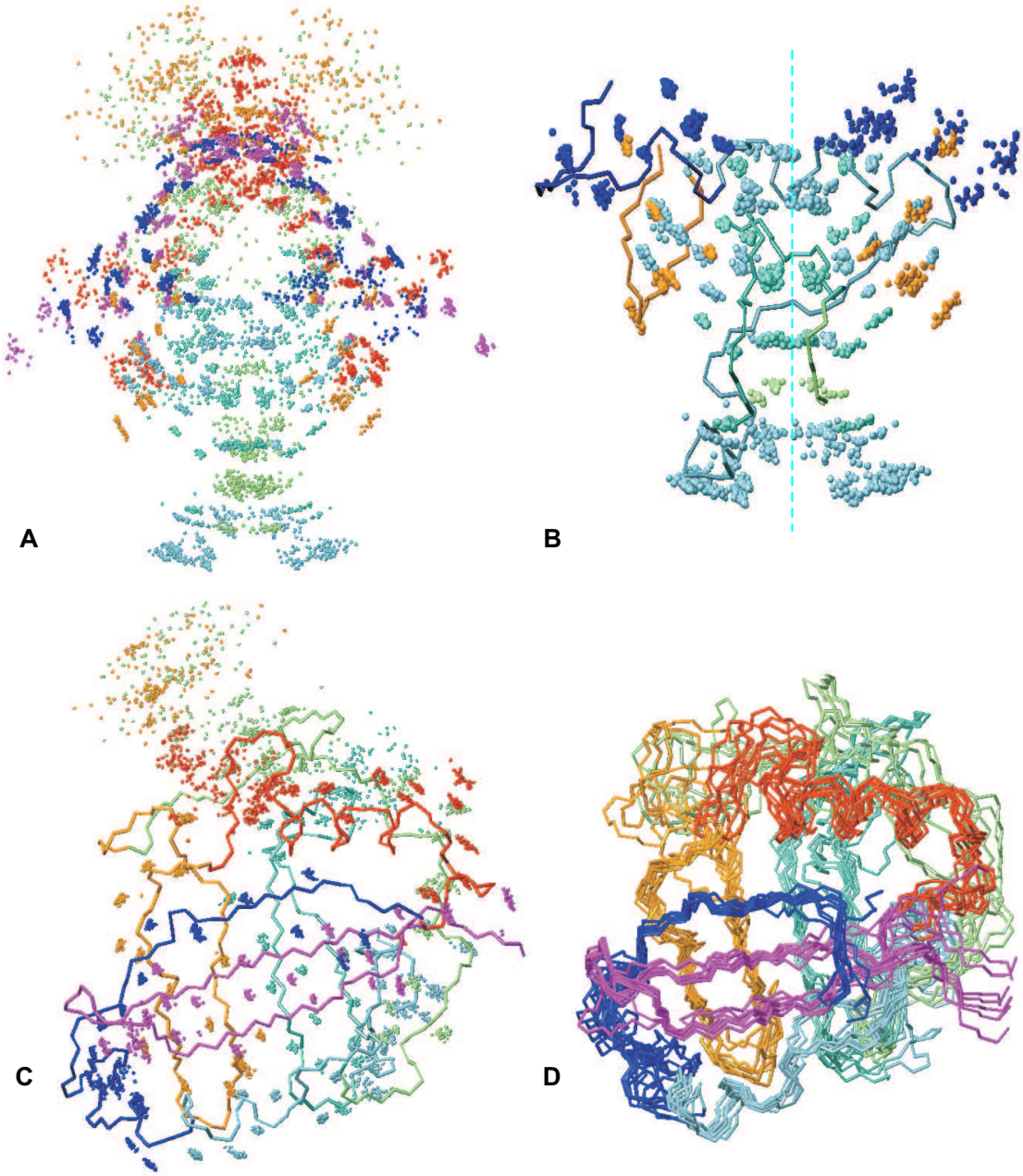


Figure 4.25: In the absence of any force-fi eld terms other than distance restraints the handedness or chirality of structures is undefi ned and two $global$ equivalent symmetry related clusters of solutions with identical low energy exist, which exhibit an inverted overall fold (A,B). In addition, $local$ mirror images can be found for α-helical segments. Without any other restraints as for NOE distances and repulsive van der Waal both $global$ subsets cannot be further classifi ed at this stage. Nevertheless, atomic distances extracted from such structures can be used to drive the sequential assignments using minimal triple resonance data. Once the sequential shift assignment becomes available, the correct fold (C,D) is inferred applying a more complete full atom force fi eld including terms for bonds, torsion and improper dihedrals, whereas key side chain NOEs are particularly important for $global$ fold discrimination.

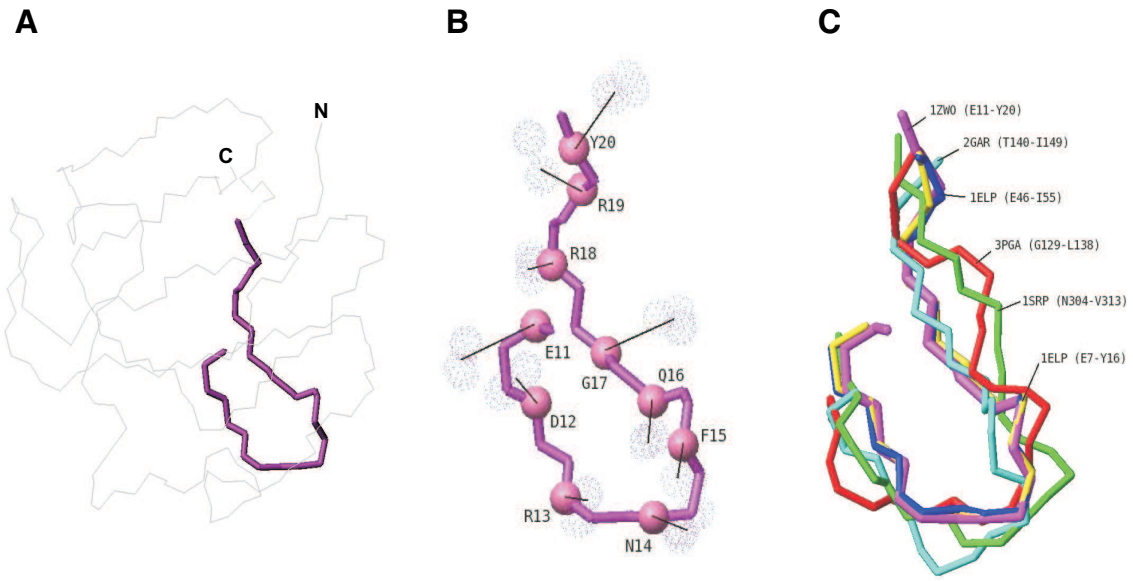


Figure 4.26: Use of atom density 'clouds' in combination with database mining and Molecular Fragment Replacement. (A) N-terminal domain of γS-crystallin (1ZWO) and location of a reverse β-turn (residue Glu11 to residue Tyr20) whose structure could not uniquely be determined by chemical shift and dipolar coupling homology search alone. (B) Resulting atomic distributions superimposed onto the reference solution structure (1ZWO). (C) Matching fragments derived from a protein data bank search using the averaged atom positions of (B) as query data.

assignment routine MONTE [63] (see section 2.5.1). As only inter-atomic distances are needed either possible global 'cloud' stereoisomer can be used in the model driven assignment process and the selection of the correct stereoisomer can be delayed until later.

To validate the value of a structural 'cloud' model in such process different input scenarios were tested. First, idealized, i.e. complete and error-free spinsystems of intra and inter-residual backbone $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts were assumed. The assignment was performed using the NOEs database with and without a atomic density cloud model.

The ideal assignment scenario is presented by the availability of a complete set of protein backbone $^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and $^{13}C'$ chemical shifts accommodated in spinsystems unimpaired by any spectral artifacts and of ideal precise chemical shift coordinates. Employing this perfect input in context with $^{1}H^{N}$-$^{1}H^{N}$ NOEs it turned out that an essential complete and unambiguous sequential assignment of the spinsystems can be achieved and the effect of improvement in terms of accuracy and extent of the sequential assignment by additionally use of a 'cloud' structure, if any, is almost negligible.

From the obtained results one can expect if a nearly completive data set of $^{13}C^{\alpha}$ *and* $^{13}C^{\beta}$ chemical shifts is disposable (especially in spinsystems with good dispersion of its $^{1}H^{N}$ and ^{15}N chemical shifts) the standard scheme for sequential backbone assignment is the best economic

choice.

From there we could conclude that, whenever possible, under favorable circumstances it is still preferable to follow the standard protocol for sequential backbone assignment using a full set of triple-resonance experiments.

However, the situation is different, when for reasons of fast relaxation the sensitivity of triple resonance experiments is severely compromised and at best only 3D HNCO, 3D HNCA and 3D HN(CO)CA experiments can be conducted with sufficient quality.

Typically, sequential backbone assignment, using $^{13}C^{\alpha}$ shifts only, while possible in theory, is usually precluded in practice due to the large $^{13}C^{\alpha}$ shift degeneracy. This is reflected in the extent of achievable sequential backbone assignment, which, while still tentatively correct (50-80% were correct), becomes very instable (only 0-25% of the sequential backbone assignment can be unambiguously made with confidence).

In conjunction with proton density structure the task becomes feasible. Inter-atomic distances thereof facilitate powerful scoring and identification of the correct long range NOES contacts in the assignment process, through with a fragment of connected spinsystem can be located much more indisputable to a distinct sequence position. In the absence of of a user provided structure model, the program generate an elongated structure template for evaluation of sequential $^{1}H^{N}$-$^{1}H^{N}$ NOE contacts to increase ambiguousness in making sequential connections of the individual spinsystems. Although sequential $^{1}H^{N}$-$^{1}H^{N}$ NOEs are helpful in restricting the spinsystems connecting possibilities, they are not sufficient for adequate resolving problems arising from $^{13}C^{\alpha}$ chemical shift degeneracy and sequential positioning of the fragments.

The using of a structural model, e.g. a homology model or in this case a proton density structure, the mid to long-range $^{1}H^{N}$-$^{1}H^{N}$ NOEs contacts interpreted in the context of the model improve the achievable reliable assignments back to 7097%. The application of the 'cloud' structure additionally circumvent putative sequence-structure alignment problems when working with a homology model.

Another problem arising from the absence of other then the more readily obtainable $^{13}C^{\alpha}$ chemical shifts for sequential linking is that (in contrast to $^{13}C^{\beta}$) with the exception of glycine, alanine, valine, isoleucine, and threonine $^{13}C^{\alpha}$ chemical shifts they do not contain a lot of residue type specific information, which further complicates their correct placement on the primary sequence of the protein. This was also reflected in the test cases where even long stretches of correctly linked spin systems failed to be unambiguously mapped onto the amino-acid sequence, which results in a low reliability of the assignments (see table 4.20).

The usage of residue specific labeling technique should counteract that weakness by providing pivotal points for the sequential assignment process. This can be realized by either labeling the sample with specific ^{15}N-labeled amino acids [135] or methods that identify the methyl containing residues [183] or by incorporation of a single residue type solely ^{13}C-labeled on its carbonyl position in an otherwise uniformly ^{15}N labeled protein. Such labeling scheme allows the acquisition of an edited HSQC spectrum in which only peaks from spinsystems that follows the ^{13}C'-labeled sequence positions are visible [113]. The residue type information than refers to the residue $i-1$ that precedes to a particular $^{1}\mathrm{H}_i^N/^{15}\mathrm{N}_i$ chemical shift pair. Both, labeling protocols that relate a $^{1}\mathrm{H}_i^N/^{15}\mathrm{N}_i$ spinsystem to a particular residue or to a residue adjacent to the $^{1}\mathrm{H}_i^N/^{15}\mathrm{N}_i$ spinsystem are employed for testing in this work. The $^{1}\mathrm{H}_i^N/^{15}\mathrm{N}_i$ of the residue types alanine and glycine were assumed to be identifiable in a straightforward manner via their characteristic $^{13}\mathrm{C}^{\alpha}$ chemical shifts and from the strong *intra* $\mathrm{C}^{\alpha} \leftrightarrow \mathrm{CH}_3^{\beta}$ NOE correlation of alanines in the ^{13}C-edited NOESY-HSQC spectrum.

The benefit of residue specific labeling for sequential assignment tests here was employed for the methyl bearing residue types valine, leucine and isoleucine, for which excellent working labeling schemes were already published (see section 2.5.2). In assignment test calculations performed in the complete absence of any triple resonance data (see below) aromatic residue type specific information was additionally applied for residue type phenylalanine and tyrosine, whereas both were introduced as being mutually commutable, e.g. a distinct $^{1}\mathrm{H}_i^N/^{15}\mathrm{N}_i$ chemical shift pair was ambiguously related to that residue types.

As indicated by the assignment results given in table 4.20, the usage of specific labeling for residue types valine, leucine and isoleucine facilitates the completing information to $^{13}\mathrm{C}^{\alpha}$ chemical shifts for obtaining an essentially complete and correct assignment of high reproducibility (e.g. confidence) when using an idealized error-free input data.

Table 4.20: Extent of achieved automated sequential assignments.

Thioredoxin			
input C_α,C_β chemical shifts, unambiguous H^N-H^N NOEs, residue type chemical shift matching			
specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
none	100/1 (100/1) [99.0/1.0 (99.0/1.0)]	39/0 (39/0)	27/0 (27/0)
(A,G,I,L,V)$_i$	101/0 (101/0) [100.0/0.0 (100.0/0.0)]	39/0 (39/0)	27/0 (27/0)
input C_α chemical shifts, unambiguous H^N-H^N NOEs, residue type chemical shift matching			
specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
none	100/1 (13/0) [99.0/1.0 (12.9/0.0)]	39/0 (0/0)	27/0 (11/0)
(A,G,I,L,V)$_i$	100/1 (100/0) [99.0/1.0 (99.0/0.0)]	39/0 (39/0)	27/0 (27/0)
input C_α chemical shifts, unambiguous H^N-H^N NOEs, residue type chemical shift matching, H^N-Methyl(I,L,V)-cloudpdb			
specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
none	100/0 (100/0) [100.0/0.0 (100.0/0.0)]	39/0 (39/0)	27/0 (27/0)
input unambiguous H^N-H^N NOEs, residue type chemical shift matching, H^N-Methyl(I,L,V)-cloudpdb			
specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
(A,G,I,L,V)$_i$	91/9 (65/0) [91.0/0.0 (65.0/0.0)]	35/4 (15/0)	27/0 (27/0)
(A,G,I,L,V,F/Y,W)$_i$	90/10 (89/5) [90.0/0.0 (89.0/0.0)]	32/7 (32/4)	27/0 (27/0)
Calmodulin			
input C_α chemical shifts, unambiguous H^N-H^N NOEs, residue type chemical shift matching			
specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
none	143/0 (0/0) [100.0/0.0 (0.0/0.0)]	87/0 (0/0)	4/0 (0/0)
(A,G,I,L,V)$_i$	145/0 (140/0) [100.0/0.0 (96.5/0.0)]	87/0 (84/0)	4/0 (4/0)
input C_α chemical shifts, unambiguous H^N-H^N NOEs, residue type chemical shift matching, H^N-Methyl(I,L,V)-cloudpdb			
specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
none	93/51 (0/0) [64.6/5.4 (0.0/0.0)]	44/43 (0/0)	4/0 (0/0)
(A,G,I,L,V)$_i$	144/0 (143/0) [100.0/0.0 (99.3/0.0)]	87/0 (87/0)	4/0 (4/0)
(A,G,I,L,V)$_{i,i-1}$	144/0 (144/0) [100.0/0.0 (100.0/0.0)]	87/0 (87/0)	4/0 (4/0)

input
C_α chemical shifts, unambiguous H^N-H^N NOEs, residue type chemical shift matching, elongated overall-α-helix

specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
none	74/68 (0/0) [52.1/7.9 (0.0/0.0)]	39/47 (0/0)	3/0 (0/0)
$(A,G,I,L,V)_i$	144/0 (115/0) [100.0/0.0 (79.9/0.0)]	87/0 (63/0)	4/0 (4/0)
$(A,G,I,L,V)_{i,i-1}$	144/0 (144/0) [100.0/0.0 (100.0/0.0)]	87/0 (87/0)	4/0 (4/0)

input
unambiguous H^N-H^N NOEs, residue type chemical shift matching, H^N-Methyl(I,L,V)-cloudpdb

specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
$(A,G,I,L,V)_i$	58/86 (0/0) [40.3/9.7 (0.0/0.0)]	22/65 (0/0)	2/2 (0/0)
$(A,G,I,L,V,F/Y,W)_i$	100/44 (2/0) [69.4/0.6 (1.4/0.0)]	48/39 (0/0)	4/0 (0/0)

input
unambiguous H^N-H^N NOEs, residue type chemical shift matching, elongated overall-α-helix

specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
$(A,G,I,L,V)_i$	2/132 (0/0) [1.5/8.5 (0.0/0.0)]	1/79 (0/0)	0/4 (0/0)
$(A,G,I,L,V,F/Y,W)_i$	84/60 (0/0) [58.3/1.7 (0.0/0.0)]	48/39 (0/0)	1/3 (0/0)

Plastocyanin

input
C_α,C_β chemical shifts, unambiguous H^N-H^N NOEs, residue type chemical shift matching

specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
none	92/1 (92/1) [98.9/0.1 (98.9/0.1)]	0/0 (0/0)	38/1 (38/1)
$(A,G,I,L,V)_i$	93/0 (93/0) [100.0/0.0 (100.0/0.0)]	0/0 (0/0)	39/0 (39/0)

input
C_α chemical shifts, unambiguous H^N-H^N NOEs, residue type chemical shift matching

specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
none	92/1 (65/0) [98.9/0.1 (69.9/0.0)]	0/0 (0/0)	38/1 (24/0)
$(A,G,I,L,V)_i$	93/0 (92/0) [100.0/0.0 (98.9/0.0)]	0/0 (0/0)	39/0 (38/0)
$(A,G,I,L,V)_{i,i-1}$	93/0 (93/0) [100.0/0.0 (100.0/0.0)]	0/0 (0/0)	39/0 (39/0)

input
C_α chemical shifts, unambiguous H^N-H^N NOEs, residue type chemical shift matching, H^N-Methyl(I,L,V)-cloudpdb

specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
none	93/0 (93/0) [100.0/0.0 (100.0/0.0)]	0/0 (0/0)	39/0 (39/0)
$(A,G,I,L,V)_i$	93/0 (93/0) [100.0/0.0 (100.0/0.0)]	0/0 (0/0)	39/0 (39/0)

input
unambiguous H^N-H^N NOEs, residue type chemical shift matching,
H^N-Methyl(I,L,V)-cloudpdb

specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
$(A,G,I,L,V)_i$	86/6 (46/0) [93.5/0.5 (50.0/0.0)]	0/0 (0/0)	38/0 (22/0)
$(A,G,I,L,V,F/Y,W)_i$	83/9 (81/5) [90.2/0.8 (88.0/0.4)]	0/0 (0/0)	36/2 (36/0)

ICln

input
C_α,C_β chemical shifts, unambiguous H^N-H^N NOEs, residue type chemical shift matching

specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
none	96/14 (93/0) [87.3/2.7 (84.5/0.0)]	14/0 (14/0)	31/1 (31/0)
$(A,G,I,L,V)_i$	108/2 (107/1) [98.2/0.8 (97.3/0.9)]	14/0 (14/0)	33/0 (33/0)
$(A,G,I,L,V)_{i,i-1}$	109/1 (108/1) [99.1/0.9 (98.2/0.9)]	14/0 (14/0)	33/0 (33/0)
$(A,G,I,L,V,F/Y,W)_i$	108/2 (107/1) [98.2/0.8 (97.3/0.9)]	14/0 (14/0)	33/0 (33/0)
$(A,G,I,L,V,F/Y,W)_{i,i-1}$	109/1 (108/1) [99.1/0.9 (98.2/0.9)]	14/0 (14/0)	33/0 (33/0)

input
C_α chemical shifts, unambiguous H^N-H^N NOEs, residue type chemical shift matching

specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
none	57/53 (0/0) [51.8/8.2 (0.0/0.0)]	0/9 (0/0)	25/5 (0/0)
$(A,G,I,L,V)_i$	108/2 (98/0) [98.2/0.8 (89.1/0.0)]	14/0 (14/0)	33/0 (33/0)
$(A,G,I,L,V)_{i,i-1}$	107/3 (99/1) [97.3/0.7 (90.0/0.9)]	14/0 (14/0)	33/0 (33/0)
$(A,G,I,L,V,F/Y,W)_i$	107/3 (105/1) [97.3/0.7 (95.5/0.9)]	14/0 (14/0)	33/0 (33/0)
$(A,G,I,L,V,F/Y,W)_{i,i-1}$	107/3 (104/1) [97.3/0.7 (94.5/0.9)]	14/0 (14/0)	33/0 (33/0)

input
C_α chemical shifts, unambiguous H^N-H^N NOEs, residue type chemical shift matching,
H^NH_α-Methyl(I,L,V,A,T)-$H_{aromatic}$-NH_2(Q,N)-cloudpdb

specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
none	80/30 (78/3) [72.7/7.3 (70.9/0.7)]	0/15 (0/0)	31/3 (31/3)
$(A,G,I,L,V)_i$	106/4 (105/1) [96.4/3.6 (95.5/0.9)]	14/0 (14/0)	33/0 (33/0)
$(A,G,I,L,V)_{i,i-1}$	107/3 (105/1) [97.3/0.7 (95.5/0.9)]	14/0 (14/0)	33/0 (33/0)
$(A,G,I,L,V,F/Y,W)_i$	108/2 (105/1) [98.2/0.8 (95.5/0.9)]	14/0 (14/0)	33/0 (33/0)
$(A,G,I,L,V,F/Y,W)_{i,i-1}$	107/3 (107/1) [97.3/0.7 (97.3/0.9)]	14/0 (14/0)	33/0 (33/0)

input
unambiguous H^N-H^N NOEs, residue type chemical shift matching,
H^NH_α-Methyl(I,L,V,A,T)-$H_{aromatic}$-NH_2(Q,N)-cloudpdb

specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
$(A,G,I,L,V)_i$	82/28 (56/0) [74.5/5.5 (50.9/0.0)]	4/10 (0/0)	31/1 (30/0)
$(A,G,I,L,V,F/Y,W)_i$	89/21 (64/0) [80.9/9.1 (58.2/0.0)]	11/3 (0/0)	32/0 (30/0)

LIM1(CRP2)

input
C_α,C_β chemical shifts, unambiguous H^N-H^N NOEs, residue type chemical shift matching

specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
none	66/4 (64/2) [94.3/0.7 (91.4/0.9)]	8/0 (8/0)	20/1 (20/0)
$(A,G,I,L,V)_i$	68/2 (66/2) [97.1/0.9 (94.3/0.9)]	8/0 (8/0)	20/0 (20/0)
$(A,G,I,L,V)_{i,i-1}$	70/0 (70/0) [100.0/0.0 (100.0/0.0)]	8/0 (8/0)	21/0 (21/0)
$(A,G,I,L,V,F/Y,W)_i$	67/3 (66/2) [95.7/0.3 (94.3/0.9)]	8/0 (8/0)	20/0 (20/0)
$(A,G,I,L,V,F/Y,W)_{i,i-1}$	70/0 (70/0) [100.0/0.0 (100.0/0.0)]	8/0 (8/0)	21/0 (21/0)

input
C_α chemical shifts, unambiguous H^N-H^N NOEs, residue type chemical shift matching

specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
none	60/10 (18/2) [85.7/4.3 (25.7/0.9)]	8/0 (0/0)	20/0 (7/0)
$(A,G,I,L,V)_i$	62/8 (62/7) [88.6/1.4 (88.6/0.0)]	8/0 (8/0)	21/0 (21/0)
$(A,G,I,L,V)_{i,i-1}$	64/6 (63/6) [91.4/0.6 (90.0/0.6)]	8/0 (8/0)	21/0 (21/0)
$(A,G,I,L,V,F/Y,W)_i$	68/2 (65/0) [97.1/0.9 (92.9/0.0)]	8/0 (8/0)	21/0 (20/0)
$(A,G,I,L,V,F/Y,W)_{i,i-1}$	70/0 (70/0) [100.0/0.0 (100.0/0.0)]	8/0 (8/0)	21/0 (21/0)

input
C_α chemical shifts, unambiguous H^N-H^N NOEs, residue type chemical shift matching, H^NH_α-Methyl(I,L,V,A,T)-$H_{aromatic}$-NH_2(Q,N)-cloudpdb

specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
none	66/4 (60/4) [94.3/0.7 (85.7/0.7)]	8/0 (8/0)	20/1 (20/1)
$(A,G,I,L,V)_i$	61/9 (58/1) [87.1/2.9 (82.9/0.4)]	8/0 (8/0)	20/0 (20/0)
$(A,G,I,L,V)_{i,i-1}$	70/0 (64/0) [100.0/0.0 (91.4/0.0)]	8/0 (8/0)	21/0 (21/0)
$(A,G,I,L,V,F/Y,W)_i$	66/4 (65/1) [94.3/0.7 (92.9/0.4)]	8/0 (8/0)	20/1 (20/0)
$(A,G,I,L,V,F/Y,W)_{i,i-1}$	70/0 (70/0) [100.0/0.0 (100.0/0.0)]	8/0 (8/0)	21/0 (21/0)

input
unambiguous H^N-H^N NOEs, residue type chemical shift matching, H^NH_α-Methyl(I,L,V,A,T)-$H_{aromatic}$-NH_2(Q,N)-cloudpdb

specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
$(A,G,I,L,V)_i$	42/26 (14/0) [61.8/8.2 (20.6/0.0)]	5/3 (0/0)	12/9 (5/0)
$(A,G,I,L,V,F/Y,W)_i$	47/23 (31/0) [67.1/2.9 (44.3/0.0)]	3/4 (3/0)	15/5 (12/0)

Cyclophilin D

input
C_α,C_β chemical shifts, unambiguous H^N-H^N NOEs, residue type chemical shift matching

specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
none	154/1 (154/1) [99.4/0.6 (99.4/0.6)]	21/0 (21/0)	50/0 (50/0)

input			
C_α chemical shifts, unambiguous H^N-H^N NOEs, residue type chemical shift matching			
specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
none	124/30 (23/0) [80.5/9.5 (14.9/0.0)]	21/0 (1/0)	38/11 (2/0)
$(A,G,I,L,V)_i$	153/2 (153/0) [98.7/0.3 (98.7/0.0)]	21/0 (21/0)	49/0 (49/0)

input			
C_α chemical shifts, unambiguous H^N-H^N NOEs, residue type chemical shift matching, H^N-Methyl(I,L,V)-cloudpdb			
specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
none	154/0 (150/0) [100.0/0.0 (97.4/0.0)]	21/0 (20/0)	50/0 (50/0)

input			
unambiguous H^N-H^N NOEs, residue type chemical shift matching, H^N-Methyl(I,L,V)-cloudpdb			
specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
$(A,G,I,L,V)_i$	136/15 (48/0) [90.1/0.9 (31.8/0.0)]	21/0 (2/0)	47/3 (30/0)
$(A,G,I,L,V,F/Y,W)_i$	140/12 (99/1) [92.1/0.9 (65.1/0.7)]	21/0 (4/0)	46/4 (42/0)

input			
C_α chemical shifts, unambiguous H^N-H^N NOEs, residue type chemical shift matching, H^N-Methyl(I,L,V,A,T)-$H_{aromatic}$-NH_2(Q,N)-cloudpdb			
specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
none	154/0 (150/0) [100.0/0.0 (97.4/0.0)]	21/0 (21/0)	50/0 (50/0)
$(A,G,I,L,V)_i$	154/1 (154/1) [99.3/0.7 (99.3/0.7)]	21/0 (21/0)	50/0 (50/0)
$(A,G,I,L,V)_{i,i-1}$	153/2 (153/1) [98.7/0.3 (98.7/0.7)]	21/0 (21/0)	50/0 (50/0)
$(A,G,I,L,V,F/Y,W)_i$	153/2 (121/0) [98.7/0.3 (78.1/0.0)]	21/0 (7/0)	50/0 (44/0)
$(A,G,I,L,V,F/Y,W)_{i,i-1}$	153/2 (153/0) [98.7/0.3 (98.7/0.0)]	21/0 (21/0)	50/0 (50/0)

input			
unambiguous H^N-H^N NOEs, residue type chemical shift matching, H^N-Methyl(I,L,V,A,T)-$H_{aromatic}$-NH_2(Q,N)-cloudpdb			
specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
$(A,G,I,L,V)_i$	125/29 (89/2) [81.2/8.8 (57.8/0.3)]	19/2 (14/2)	47/3 (39/0)
$(A,G,I,L,V,F/Y,W)_i$	133/19 (103/3) [87.5/2.5 (67.8/1.0)]	19/2 (15/2)	45/5 (40/0)

input			
C_α chemical shifts, unambiguous H^N-H^N NOEs, residue type chemical shift matching, H^NH_α-Methyl(I,L,V,A,T)-$H_{aromatic}$-NH_2(Q,N)-cloudpdb			
specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
none	154/0 (150/0) [100.0/0.0 (97.4/0.0)]	21/0 (21/0)	50/0 (50/0)
$(A,G,I,L,V)_i$	154/1 (154/1) [99.3/0.7 (99.3/0.7)]	21/0 (21/0)	50/0 (50/0)
$(A,G,I,L,V)_{i,i-1}$	154/1 (154/1) [99.3/0.7 (99.3/0.7)]	21/0 (21/0)	50/0 (50/0)
$(A,G,I,L,V,F/Y,W)_i$	154/1 (154/1) [99.3/0.7 (99.3/0.7)]	21/0 (21/0)	50/0 (50/0)
$(A,G,I,L,V,F/Y,W)_{i,i-1}$	154/1 (154/1) [99.3/0.7 (99.3/0.7)]	21/0 (21/0)	50/0 (50/0)

input
unambiguous H^N-H^N NOEs, residue type chemical shift matching,
$H^N H_\alpha$-Methyl(I,L,V,A,T)-$H_{aromatic}$-NH_2(Q,N)-cloudpdb

specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
$(A,G,I,L,V)_i$	125/29 (89/2) [81.2/8.8 (57.8/0.3)]	19/2 (14/2)	47/3 (39/0)
$(A,G,I,L,V,F/Y,W)_i$	133/19 (103/3) [87.5/2.5 (67.8/1.0)]	19/2 (15/2)	45/5 (40/0)

Human Carbonic Anhydrase I

input
C_α,C_β chemical shifts, unambiguous H^N-H^N NOEs, residue type chemical shift matching

specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
none	241/1 (219/0) [99.6/0.4 (90.5/0.0)]	18/0 (17/0)	76/0 (69/0)
$(A,G,I,L,V)_i$	241/0 (238/0) [100.0/0.0 (98.8/0.0)]	18/0 (18/0)	76/0 (74/0)
$(A,G,I,L,V)_{i,i-1}$	241/0 (241/0) [100.0/0.0 (100.0/0.0)]	18/0 (18/0)	76/0 (76/0)

input
C_α chemical shifts, unambiguous H^N-H^N NOEs, residue type chemical shift matching

specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
none	162/75 (2/0) [68.4/31.7 (0.8/0.0)]	18/0 (10/0)	76/0 (61/0)
$(A,G,I,L,V)_i$	241/1 (185/0) [99.6/0.4 (76.8/0.0)]	18/0 (10/0)	76/0 (61/0)
$(A,G,I,L,V)_{i,i-1}$	241/1 (222/0) [99.6/0.4 (91.7/0.0)]	18/0 (17/0)	76/0 (74/0)

input
C_α chemical shifts, unambiguous H^N-H^N NOEs, residue type chemical shift matching,
H^N-Methyl(I,L,V)-cloudpdb

specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
none	73/163 (5/0) [30.9/9.1 (2.1/0.0)]	0/17 (0/0)	25/48 (1/0)
$(A,G,I,L,V)_i$	234/6 (234/0) [97.5/0.5 (97.5/0.0)]	18/0 (18/0)	76/0 (76/0)
$(A,G,I,L,V)_{i,i-1}$	239/2 (239/0) [99.2/0.8 (99.2/0.0)]	18/0 (18/0)	76/0 (76/0)

input
unambiguous H^N-H^N NOEs, residue type chemical shift matching,
H^N-Methyl(I,L,V)-cloudpdb

specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
$(A,G,I,L,V)_i$	185/53 (13/0) [77.7/2.3 (5.5/0.0)]	0/17 (0/0)	73/3 (12/0)
$(A,G,I,L,V,F/Y,W)_i$	207/31 (187/5) [87.0/13.0 (78.6/2.1)]	12/6 (8/1)	75/0 (73/0)

To evaluate the assignment success in cases where only a more realistic input data is available the effect of some imperfections were investigated using the experimental data set of cyclophilin D (table 4.21). First, the influence of limited peak picking accuracy was investigated.

Typically the precision at which a peak position can be determined depends on both the signal to noise ratio (S/N) and the line width at half height ($LW_{1/2}$). It can be expressed as $LW_{1/2}/(S/N)$. Assuming a typically S/N ratio of 20 and a line width determined by an acqui-

sition time in the $^{13}C^{\alpha}$ dimension of 7 ms, spectral noise was introduced into the input data by a adding a value randomly chosen within a range of $\pm$ 0.2 ppm to each *intra* $^{13}C_i^{\alpha}$ and *inter* $^{13}C_{i-1}^{\alpha}$ chemical shift entry.

As a result the extent of obtained correct assignment using solely these $^{13}C_i^{\alpha}$/$^{13}C_{i-1}^{\alpha}$ chemical shift database and $^{1}H^{N}$-$^{1}H^{N}$ NOEs contacts for sequential linking dropped down by a factor of two to 45% (whereas only 6.2% could be made with confidence) and more than 50% were tentatively wrong. The additionally usage of specifc labeling for residue types leucine, isoleucine, and valine ('ILV-labeling') enables an improvement to 72% correct assigments (50% found as reliable). Even in the presence of more residue specific assignment restraints arising from accessorily adding labeling information that refers to residues precceding to a particular glycine, alanine, leucine, isoleucine, or valine residue type (^{13}C' labeling) the achieved correct and reliable assignment extent do not exceed 65%, whereas about 17% incorrect assignments were remaining.

The success rates of sequential backbone assignments were completely different when using a minimalistic 'cloud' structure generated from $^{1}H^{N}$/CH_3^{ILV} NOE distance restraints only. Omitting any further aid from specific labeling 97.3% correct assignments were obtained (thereof 64.2% reliable). Furthermore, an essentially complete reliable correct assignment (96%) was accomplished by additional implying of specific labeling for residue types leucine, isoleucine, and valine. An absolute reliable and correct assignment was found for spinsystems related to residues in regulary secondary structure elements.

In a next step, the quality of the *intra* $^{13}C_i^{\alpha}$ and *inter* $^{13}C_{i-1}^{\alpha}$ chemical shift database was decreased by removing any 96 sequential *inter* $^{13}C_{i-1}^{\alpha}$ chemical shift belonging to residues located outside of secondary structure elements (condition B). As given in table ?, the only usable and essentially error-free sequential assignment results were obtained when applying a 'cloud' structure model in context with 'ILV-labeling' for the assignment process.

Then, in addition 25 selected *intra* $^{13}C_i^{\alpha}$ chemical shifts related to residues found in bend regions in the native fold of the protein were deleted from the input. Whereas only 17% correct assignment were observed using $^{13}C_i^{\alpha}$ chemical shifts and backbone amide $^{1}H^{N}$-$^{1}H^{N}$ NOE distances only, 92.2% of the spinsystems were found to be error-free assigned. The reliable part of the sequential assignments belonging to residues located in α or β-regions of cyclophilin D was found to be essentially complete (97.1%) and starkly correct.

Table 4.21: Effects of limited precision of chemical shifts and completeness of chemical shift database[a].

cyclophilin D using sparse chemical shift data $\pm$ H^N-CH_3(I,L,V)-cloudstructure			
input C_α chemical shifts, unambiguous H^N-H^N NOEs, no residue type labeling			
extent of chemical shifts	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
condition A	66/78 (9/0) [45.8/4.2 (6.2/0.0)]	11/8 (0/0)	11/32 (0/0)
condition B	44/103 (5/0) [29.9/0.1 (3.4/0.0)]	1/20 (0/0)	6/42 (0/0)
condition C	25/123 (8/0) [16.9/3.1 (5.4/0.0)]	1/20 (0/0)	1/46 (0/0)
input C_α chemical shifts, unambiguous H^N-H^N NOEs, cloudpdb, no labeling			
extent of chemical shifts	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
condition A	147/4 (97/3) [97.3/0.7 (64.2/1.0)]	19/1 (9/0)	49/0 (27/0)
condition B	109/42 (25/0) [72.2/7.8 (16.6/0.0)]	2/18 (0/0)	36/14 (0/0)
condition C	131/20 (18/0) [86.8/3.2 (11.9/0.0)]	14/7 (1/0)	49/1 (0/0)
input C_α chemical shifts, unambiguous H^N-H^N NOEs, $(A,G,I,L,V)_i$ labeling			
extent of chemical shifts	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
condition A	107/41 (75/1) [72.3/7.7 (50.7/0.7)]	19/1 (12/0)	33/14 (21/1)
condition B	79/72 (39/0) [52.3/7.7 (25.8/0.0)]	11/9 (11/0)	13/34 (1/0)
condition C	86/66 (54/0) [56.6/3.4 (35.5/0.0)]	19/2 (11/0)	11/38 (2/0)
input C_α chemical shifts, unambiguous H^N-H^N NOEs, cloudpdb, $(A,G,I,L,V)_i$ labeling			
extent of chemical shifts	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
condition A	147/5 (146/3) [96.7/0.3 (96.0/1.0)]	20/0 (20/0)	50/0 (50/0)
condition B	150/3 (123/0) [98.0/1.0 (80.4/0.0)]	18/2 (13/0)	50/0 (45/0)
condition C	141/12 (135/9) [92.2/0.8 (88.2/0.9)]	20/1 (18/0)	50/0 (46/0)
input C_α chemical shifts, unambiguous H^N-H^N NOEs, $(A,G,I,L,V)_{i,i-1}$ labeling			
extent of chemical shifts	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
condition A	124/24 (96/0) [83.8/6.2 (64.9/0.0)]	18/2 (17/0)	44/4 (33/0)
condition B	130/22 (88/0) [85.5/4.5 (57.9/0.0)]	21/0 (16/0)	38/9 (29/0)
condition C	131/21 (95/0) [86.2/3.8 (62.5/0.0)]	21/0 (16/0)	41/8 (31/0)
input C_α chem. shifts, unambiguous H^N-H^N NOEs, cloudpdb, $(A,G,I,L,V)_{i,i-1}$ labeling			
extent of chemical shifts	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
condition A	147/5 (147/5) [96.7/0.3 (96.7/0.3)]	20/0 (20/0)	50/0 (50/0)
condition B	151/3 (139/0) [98.0/0.9 (90.3/0.0)]	21/0 (20/0)	48/2 (47/0)
condition C	140/13 (136/5) [91.5/0.5 (88.9/0.3)]	20/1 (20/0)	48/2 (48/0)

[a] condition A adding random noise of $\pm$0.2 ppm to each $C^{\alpha}_{i,i-1}$.

Condition B as A, but also removing inter C^{α}_{i-1} chemical shifts of spinsystems related to residues outside of regular secondary structure.
Condition C as B, but deleting 25 C^{α}_{i} chemical shifts of residues structural located in bend regions.

Finally, the influence of substantial structural erroneous 'cloud' model in an assignment protocol was studied. Incorrect atom density 'clouds' can potentially arise from mis-assignment of NOEs. Interestingly, even the presence of several incorrect long-range NOEs expected to hardly distort the atom density model was not detectable by a higher energy term U_{NOE}, indicating the importance of covalent restraints for derivation of a 'self-consistent' NOE distance network compatible with the native fold (see figure 4.27). On the other side, erroneously assigned NOE distances related to atoms of residues located close in the primary sequence results only in rather negligible strains of the atom density models. This is because the absence of covalent connections makes the distance network more flexible and impede the propagation of such local distortions.

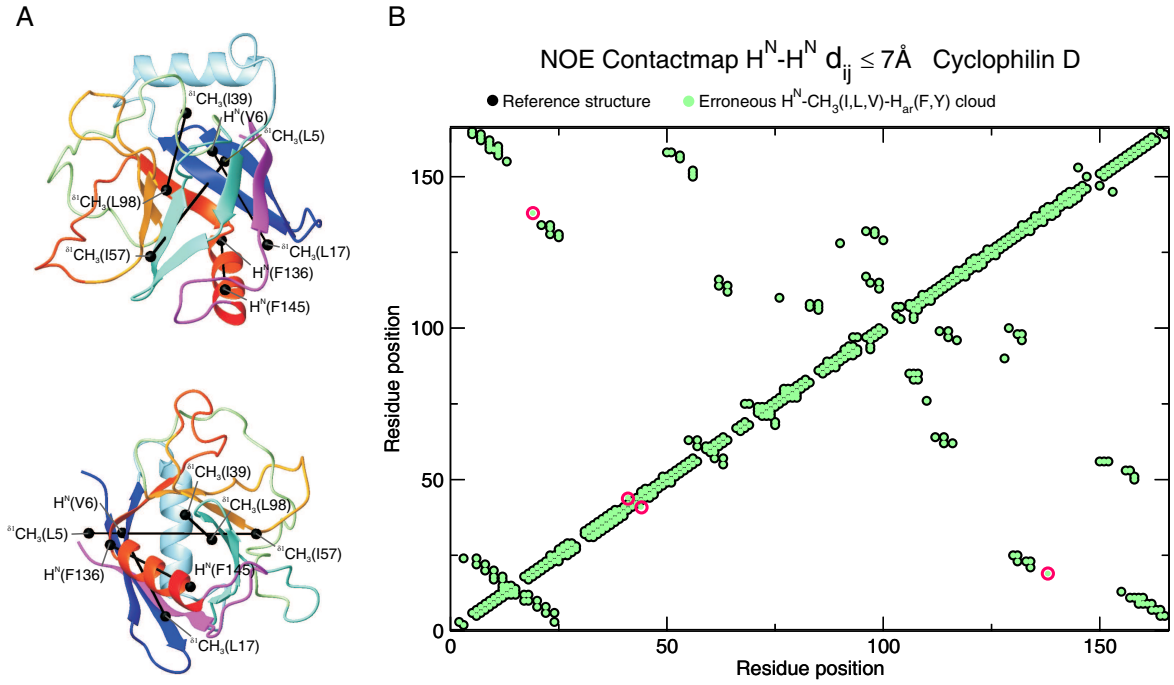


Figure 4.27: In contrast to wrong short- to medium-range NOEs (i.e. the NOE distance related protons are le 4 residues apart in the primary sequence) incorrect long-range lead to atomic density distributions considerable in disagreement with the native protein fold. (A) Four erroneous long-range NOE distance restraints were artifi cially added to the structure calculation input of cyclophilin D. (B) Interestingly, despite of their introduction the local inter-atomic environment (considering small distances within NOEs can be typically observed) remain essentially identical as shown by the backbone H^N-H^N NOE contact map, where large black circles refer to the interatomic H^N-H^N distances derived from the NMR reference structure (derived conventionally) le 7Å for which a NOE contact was present in the experimentally input database, and circles colored in cyan denotes the analog H^N-H^N atom distances inferred from the erroneous 'cloud' structure.

Therefore, for the applied test case an erroneous 'cloud' model was generated from a NOE input data set that additionally contained a selection of four incorrect long-range NOE distances (R_{e1} to R_{e4}) between pairs of atoms far apart in the correct tertiary structure. The selection was done according their critical impact on the structural results.

Table 4.22: Effects of structural errors in the atom density 'clouds' applied for sequential assignment

Cyclophilin D using faulty[a] H^N-CH_3(I,L,V)-$H_{aromatic}$(F,Y)-cloud structure in context with full, sparse (condition C) or no chemical shift database			
input all exp. C_α chem. shifts, unambiguous H^N-H^N NOEs, erroneous cloudpdb			
specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
none	154/1 (141/0) [99.3/0.7 (91.0/0.0)]	21/0 (11/0)	50/0 (50/0)
$(A,G,I,L,V)_i$	154/1 (154/1) [99.3/0.7 (99.3/0.7)]	21/0 (21/0)	50/0 (50/0)
input sparse exp. C_α chem. shifts (condition C), unambiguous H^N-H^N NOEs, erroneous cloudpdb			
specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
none	19/126 (0/0) [13.1/6.9 (0.0/0.0)]	1/19 (0/0)	9/41 (0/0)
$(A,G,I,L,V)_{i,i-1}$	134/19 (122/0) [87.6/2.4 (79.7/0.0)]	21/0 (21/0)	40/10 (39/0)
$(A,G,I,L,V,F/Y,W)_{i,i-1}$	142/8 (123/2) [94.7/0.3 (82.0/0.3)]	21/0 (18/0)	50/0 (48/0)
input unambiguous H^N-H^N NOEs, erroneous cloudpdb			
specific labeling	**# correct assignment/errors (# reliable assignment/errors)**		
	total [% total]	α-helix	β-sheet
none	3/145 (0/0) [2.0/8.0 (0.0/0.0)]	0/20 (0/0)	0/49 (0/0)
$(A,G,I,L,V)_i$	127/25 (74/5) [83.5/6.5 (48.7/0.3)]	19/2 (10/2)	40/10 (33/2)
$(A,G,I,L,V)_{i,i-1}$	139/13 (129/1) [91.5/0.6 (84.9/0.7)]	19/2 (19/0)	50/0 (47/0)
$(A,G,I,L,V,F/Y,W)_i$	131/20 (95/6) [86.8/3.2 (62.9/1.0)]	16/5 (12/2)	40/10 (37/3)
$(A,G,I,L,V,F/Y,W)_{i,i-1}$	140/13 (138/1) [91.5/0.5 (90.2/0.7)]	19/2 (19/0)	47/3 (47/0)

[a]'erroneous cloud structure', generated by adding four incorrect long range NOE to the experimental input NOE database [atom (location)]:
(1) L5-$CH_{3,\delta1}$ (facing outwards at the top of β-barrel) to I57-$CH_{3,\delta1}$ (exterior at bottom opposite site of β-barrel) [29.2Å], (2) backbone V6-H^N(begin strand β_2) to L17-$CH_{3,\delta1}$ (end of strand β_2) [19.1Å], (3) L39-$CH_{3,\delta1}$ (core of β-barrel) to L98-$CH_{3,\delta1}$ (core of β-barrel) [11.7Å], and (4) backbone M136-H^N (start α_2 helix) to F145-H^N (end α_2 helix) [15.4Å].

The incorrect restrain R_{e1} connect the N- and the C-terminus of the α_2-helix, R_{e2} the N- and the C-terminus of the β_2-strand, restraint R_{e3} the top and bottom of the central βbarrel and R_{e4} strand β_1 and β_7 along the core (figure 4.27.A). All these errors while detrimental for the resulting proton density cloud distributions per-se (which did not longer resemble the protein tertiary structure), had surprisingly little influence on the results of the model-driven assignment process as they left the overwhelming majority of native tertiary contacts intact (figure 4.27.B).

The extend of the sequential assignment obtained is only marginally lower (87.6% compared to 92.2% correct assignments) to that achieved with a correct structural model applied to the same input data.

The outlined results indicate that the derived atomic 'cloud' models enable sequential resonance assignments in cases of incomplete triple resonance chemical shift data and limited precision of the peak picking even when fatal structural errors caused by misassignment of some long-range NOEs are apparent.

It should be mentioned that these artificially introduced incorrect long-range NOE restraints can principally be identified by an unusual distribution of NOE contacts within the initial input NOE density.

Once the backbone chemical assignment has become available the sequential sidechain assignment can be obtained in a automated protocol as presented in section 2.5.3 using the H^N-CH_3 distances of the derived 'cloud' structure model as additionally constrain for that purpose. In a final step, a structure with connecting bonds can be calculated starting from a covalent template and employing the whole set of improper and chiral restraints. The presence of these covalent restraints in context with sidechain methyl key NOEs facilitates the identification of the correct stereoisomere and lead to a refined structure model, whose quality can be further increased using methods of bioinformatics.

Taking together, the presented method facilitates in a first step the generation of a (low-resolution) non-covalent representation of protein tertiary structure (without any knowledge of the sequential assignment), which can be used e.g. for rapid fold validation purposes or model-driven automated sequential assignment. In a second step a covalent structure of reasonable quality can be derived, which can be serve as initial structure template for well established automated structure refinement routines that need backbone resonance assignment as prerequisite but taking complete NOESY data (as derived from an uniformly ^{15}N,^{13}C-labeled sample) into account.

4.6 Capability of the Automated binding site mapping schedule (ROSA)

The automated binding site mapping approach described in section 2.8 has been applied to a test set of eight published proteins-ligand complexes whose backbone chemical shifts were deposited at the BioMagRes databank (BMRB) for both, the *apo* and the *holo* form. The extent of the available chemical shifts were essentially complete for all cases (see table 3.22 in section 3.6). However, not any sequential assignment was part of the program input.

The spatial distribution of the residues identified as being next or part of the binding site are given in figure 4.28. If any atom of these residues is located within a 8Å cutoff to a ligand atom the related residue is colored with a green surface, otherwise the residue surface is marked in red. To achieve an adequate graphical representation of the binding site prediction 10 to 15 of the most shifted residues (indicated as black dots in figure 4.28) were selected for the program input.

The algorithm uses the measured chemical shift differences observed in ^{1}H-^{15}N correlation spectra to identify starting points for the subsequent analysis (e.g. fragments or residue positions which are altered upon ligand binding). The residues are then attributed to the binding site with using sequential triple-resonance connectivity information obtained by triple-resonance spectroscopy and spatial proximity information derived from the 3D coordinates (see Materials and Methods).

As consequence of the program algorithm, in which these residues with most significant $^{1}H^{N}$ and ^{15}N chemical shift changes serves as starting point for subsequent analysis (mapping of connected spinsystems on the protein tertiary structure by spatial criteria), a strong correlation was found between results obtained by this method and those solely inferred from conventional chemical shift perturbation analysis. An unique binding site prediction was obtained whenever exact one spatial patch of residues with most pronounced chemical shift changes upon ligand binding was present. This is demonstrated the examples MBP + Maltotriose (Figure 3b), SNase+Thymidine 3',5'-di-P in (Figure 3c), Ap4a + ATP (Figure 3e), Biotin + CoA (Figure 3d), Adenylate Kinase + AP5A (Figure 3a), S100B + p53t (Figures 3f) and CRBPII + all trans retinol.

Table 4.23: Results of automated binding site mapping approach

Complex (Protein/) (Ligand)	**10 most shifted residues** ($\bar{x} \pm 2 \cdot \sigma_x$)[a]	**Mapped residues using most n shifted residues (close to correct BS[b] [%])**	**Equiv. to ^{1}H-^{15}N CSM [c]**
CRBPII/ *all-trans* Retinol	33, 34, 36, 56, 57, 59, 60, 61, 62, 77 (0.126 ± 0.108)	n=5: 44, 61, 62, 63, 94 (100%) n=10: 33, 34, 43, 44, 54, 59, 60, 61, 62, 63, 68 (90%)	0 16.6
Adenylate kinase/ AP5A	7, 26, 31, 32, 33, 58, 63, 64, 67, 68, (0.388 ± 0.064)	n=6: 31, 85, 86, 175 (100%) n=10: 31, 67, 85, 86, 123, 125, 126, 127, 132, 133, 134, 146 (67.8%)	16.7 10.0
Ap4a hydrolase/ ATP	14, 18, 34, 44, 46, 47, 82, 103, 147, 149 (0.291 ± 0.088)	n=5: 46, 81, 82, 103, 105 (80%) n=10: 13, 45, 46, 47, 80, 81, 82, 83, 102, 103, 105, 147 (75%)	40.0 25.0
BtnCoA/ Biotin	90, 92, 94, 96, 98, 100, 102, 116, 117, 118 (0.065 ± 0.179)	n=5: 90, 92, 102, 121, 122 (100%) n=10: 90, 92, 94, 98, 101, 102, 120, 122, 124 (88%)	60.0 55.6
MBP/ Malto-triose	110, 155, 168, 258, 259, 299, 302, 303, 332, 341 (0.195 ± 0.071)	n=5: 108, 302 (50%) n=10: 108, 110, 168, 258, 259, 299, 302, 303 (62.5%)	0 50.0
S100B/ P53t	12, 17, 18, 19, 20 21, 26, 27, 30, 31 (1.171 ± 0.076)	n=5: 42, 44, 46, 61, 67, 68 (83%) n=10: 42, 43, 46, 47, 66, 67 (67%)	
Snase/ Thymidin	37, 38, 39, 41, 54, 84, 110, 113, 114, 115 (0.152 ± 0.115)	n=5: 37, 38, 39, 91, 103, 104, 110 (43%) n=10: 23, 37, 38, 39, 40, 54, 103, 105, 110, 111, 112, 137 (42%)	57.1 33.3

[a] Residues whose absolute $|^{1}$H-^{15}N$|$ chemical difference value upon ligand binding was most signifi cant, $\bar{x}$ in brackets denotes the mean overall chemical shift change and σ its associated standard deviation. [b] BS = binding site. A mapped residue was considered next to the correct binding site whenever at least one inter-atomic distance to the ligand closer than a cutoff (12Å for mbp, 8Å for others) value were found. [c] CSM = chemical shift mapping. Percentages are the ratio of the number of residues identifi ed by both, standard ^{1}H-^{15}N chemical shift mapping and ROSA to those obtained by ROSA solely.

However, note that the obtained predictions are made without the prerequisite of any (time consuming) sequential resonance assignment or the acquisition of a set of standard triple experiments for both, the *apo* and the *holo* form. Just two 2D ^{1}H^{N}-^{15}N HSQC spectra have to be acquired for identifying and classification of the spinsystem with most obvious chemical shift changes by ligand binding.

Another situation is given in the example of SNase (figure 4.28.G), where two significant patches of shifted signals were existent in the input. Although consequently two predictions were proposed by the program, a more closer decision could be made by the location of identified residues related to very most perspicuously shifted spinsystems, which were exclusively ac-

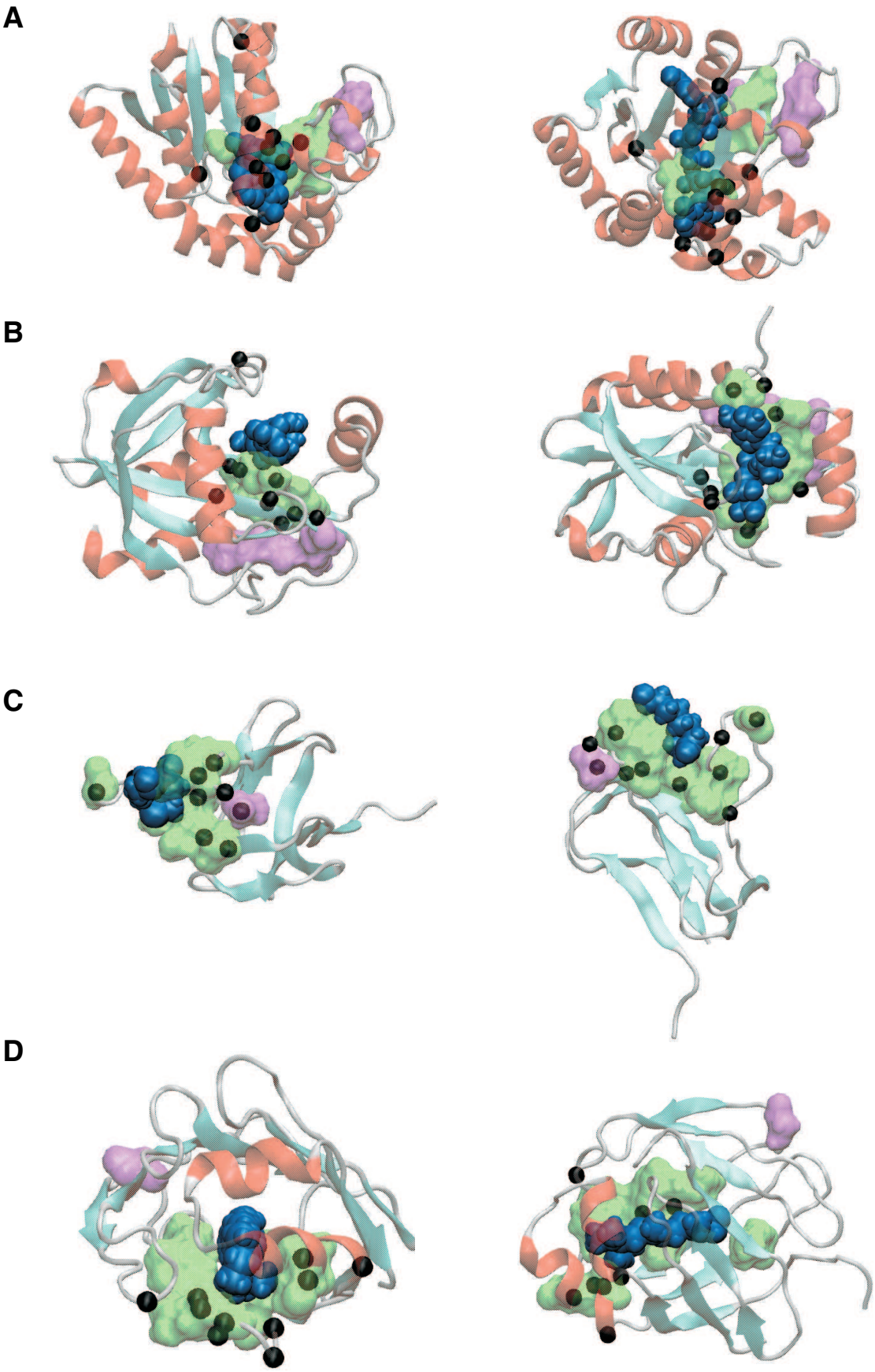


Figure 4.28: Graphical illustration of ROSA binding site mapping results shown on the $holo$ protein structures. Residues whose backbone ^{1}H and ^{15}N resonances exhibit the most signifi cant chemical shift changes are shown in black spheres. Residues correctly identifi ed to be part or next to the binding site are represented by transparent green surfaces, whereas residues suggested by ROSA but without

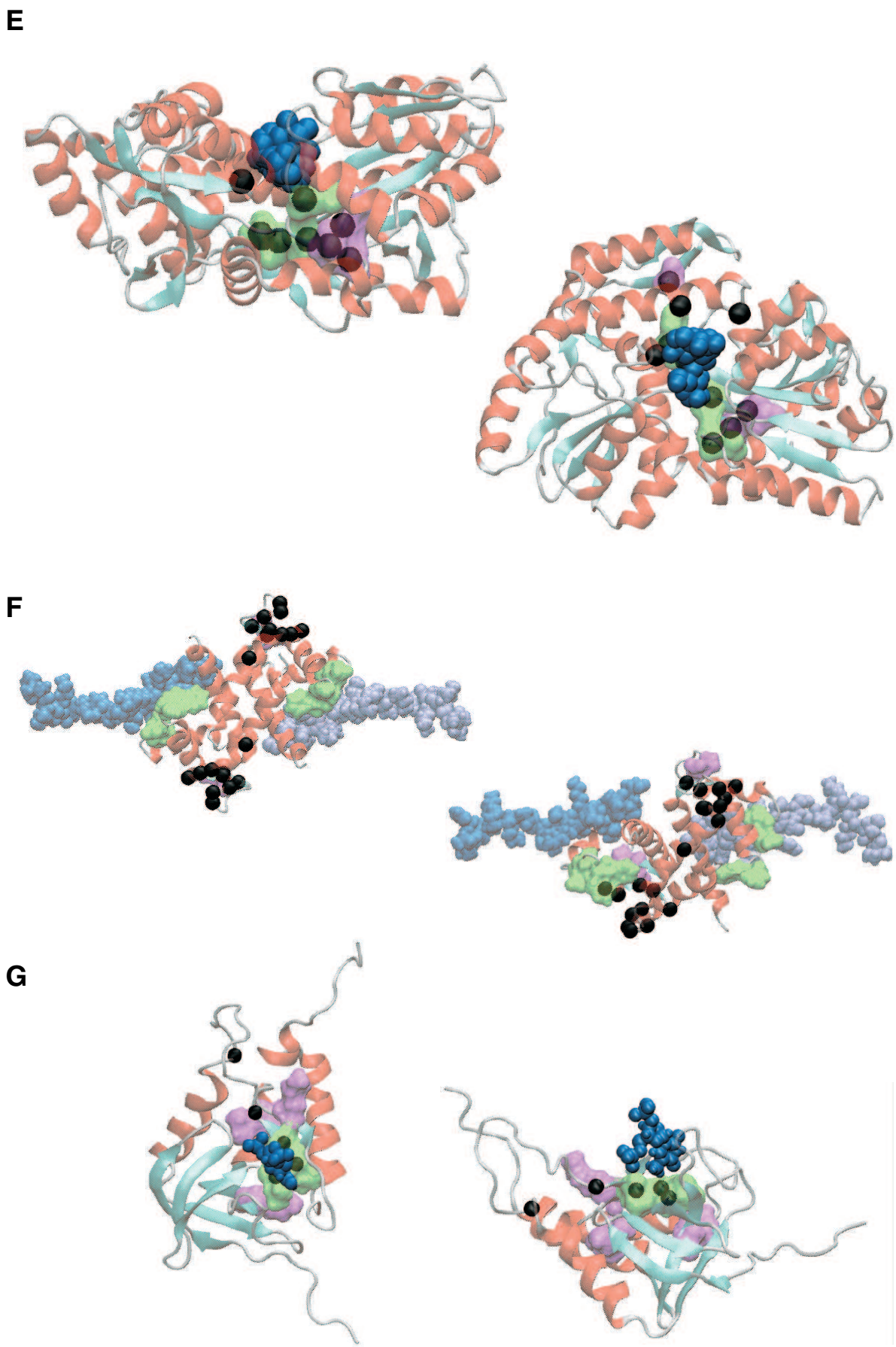


Figure 4.28: ROSA binding site mapping results continued.(E) Maltose binding protein/ Matltotrose (F) S100B/p53t, (G) Snase/Thymidin.

commodated next to the correct binding region presumably due to (close to the heterocyclic ring). Furthermore, residues in the $\beta_{6/7}$-loop of SNase (located near the residual phosphates) could not be identified as part of the binding site, despite reasonable overall chemical shift dispersion or successive alteration of the distance cutoff value. It seems that in loosely packed loop regions (with inherent low residue density) the spatial neighborhood criteria cannot be easily fulfilled adequate enough for unambiguous detection.

The presented results were generated using standard parameter values for sequential connection and a distance cutoff value of 8Å for the spatial neighborhood criteria of the connected spinsystems. These values were found to facilitate the generation of a sufficient amount of spinsystem fragments accessibly for scoring of the neighborhood criteria, whereas spinsystems related to residues affected by long-range chemical shift disturbation effects were efficiently excluded from the analysis (which is vital to get an adequate resolved localization of the binding site). As mentioned in section 2.8, and because the efficiency of the binding site prediction protocol depends on a manageable amount of sequential spinsystem connection possibilities to find unambiguous predictions, the methods works best if at least two types of backbone chemical shifts (e.g. $^{13}C^{\alpha}_{i,i-1}$, $^{13}C^{\beta}_{i,i-1}$ or $^{13}C^{\alpha}_{i,i-1}$, $^{13}C'_{i,i-1}$) are available for sequential connecting. However, an exception was the example of calmodulin for which sufficient sequential connectivity information for reliable binding site mapping was already achieved by including $^{13}C^{\alpha}_{i,i-1}$ chemical shifts only.

In summary, once a list of spinsystems and a properly structure model is available the tested approach provide fast information on putative binding site locations without apriori knowledge of sequential assignment. Therefore it's suggested to be applied at the early stage of a NMR investigation. In addition, the obtained results could be verified by alternative spectroscopically methods, such as cross-saturation techniques across the binding interface where the mapping is based on dipole-dipole interactions [37]. Furthermore, using well-established relaxation-optimized NMR observation techniques especially in combination with cold probe technology the application could also be employed for large molecular weight protein complexes.

List of Figures

1.1 Multiple alignment of ICln sequences 13
1.2 A model for the mitochondrial permeability transition pore complex 23
1.3 Multiple sequence alignment of BASP1. 27
1.4 Multiple alignment of osteopontin sequences. 29
1.5 Osteopontin in tumor progression - a flowchart. 32

2.1 Suite of standard triple resonance experiments for sequential backbone chemical shift assignment. 67
2.2 Observed peak patterns for a stretch of five residues of brain abundant protein 1 (BASP1). 68
2.3 Through bond and through space correlation experiments for sequential sidechain chemical shift assignment. 71
2.4 Prochiral assignment using $^{3}J_{C_{\gamma}N}$/$^{3}J_{C_{\gamma}C'}$ couplings. 78
2.5 Ramachandran and χ_1,χ_2 plot of cyclophilin D. 88
2.6 Conformational energy distribution derived from artificial PROSA poly-protein. . . 92
2.7 Total interaction energy as a function of primary sequence position for cyclophilin D. 93
2.8 Postscript output of the backbone assignment program MONTE. 99
2.9 Selective labeling scheme for residue types leucine, isoleucine and valine. 102
2.10 Flowchart of the applied routine for automated assignment of methyl groups. . . 105
2.11 Summary of the direct method protocol developed in this work. 111
2.12 Distance plot for Cyclophilin D and definition of the selection criteria. 115
2.13 Scheme for automated binding site prediction tool (ROSA). 120

3.1 Samples of ICln used for NMR analysis. 122
3.2 Structures of potential ICln ligands applied in titration experiments. 123

3.3 Primary sequence of *homo sapiens* (human) cyclophilin D. 125
3.4 Primary sequence of the LIM1 domain from CRP2. 126
3.5 Characteristic NOE contacts found in protein secondary structure elements. . . . 138
3.6 Scaling of the forcefield weighting factors in the applied simulating annealing protocol. 143

4.1 800 MHz 2D $^{1}H^{N}$,^{15}N HSQC of N-terminal ICln. 154
4.2 2D $^{1}H^{N}$,^{15}N HSQC overlay of full length wild type ICln with its truncated N- or C-terminal forms. 155
4.3 2D $^{1}H^{N}$,^{15}N HSQC of the C-terminal ICln. 163
4.4 Secondary chemical shift plot for (A) truncated (N-terminal) and (B) C-terminal ICln. 166
4.5 ^{15}N-Relaxation T_1 and T_2 values for N- and C-terminal ICln. 168
4.6 Structural results for recombinant N-terminal ICln. 170
4.7 Overall electrostatic potentials comparison of ICln with other PH domains. 174
4.8 Structural localization of shifted residue in ICln upon integrin peptide binding. . . 176
4.9 Lsm4 titration experiment of ICln. 179
4.10 800 MHz 2D $^{1}H^{N}$,^{15}N HSQC of cyclophilin D. 181
4.11 Structural results for wild type $homosapiens$ cyclophilin D. 191
4.12 Comparison of the derived NMR model with the published x-ray model of Cyclophilin D. 194
4.13 Secondary chemical shift plot of $quail$ LIM1/CRP2. 199
4.14 Structural results of recombinant $quail$ LIM1/CRP2. 200
4.15 2D $^{1}H^{N}$-^{15}N HSQC of $chicken$ brain abundant protein 1 (BASP1). 204
4.16 Secondary chemical shift plot of BASP1. 209
4.17 2D $^{1}H^{N}$-^{15}N HSQC of *quail* osteopontin. 210
4.18 Secondary chemical shift plot for osteopontin. 216
4.19 Experimentally derived relaxation parameters of osteopontin. 218
4.20 Spectral density function of osteopontin derived at frequencies of 0, 60.8 MHz (ω_N), and 600.1 MHz (ω_{H^N}). 219
4.21 Extent of obtained results for cyclophilin D using automated sidechain methyl group assignment. 222
4.22 Surface energy graphs of the best H^N/CH_3^{ILV} structures. 224

4.23 H^N/CH_3^{ILV} structure ensembles for ICln, cyclophilin D and lipocalin Q83. 225
4.24 Requirement of sidechain NOE distances for generation of a primarily model for subsequent structure determination. 227
4.25 Structural results obtained at several stages of the direct method using cyclophilin D as example. 236
4.26 Use of atom density 'clouds' in combination with database mining and Molecular Fragment Replacement. 237
4.27 Impact of serious erroneous long-range NOE contacts on the derived atomic density distributions. 248
4.28 Graphical illustration of ROSA binding site mapping results shown on the $holo$ protein structures. 253
4.28 ROSA binding site mapping results continued. 254

List of Tables

2.1 Structure quality indicators used by PROCHECK. 87

2.2 Empirical interpretation of the quality index for NMR structures obtained by DACA. 95

2.3 ^{1}H and ^{13}C methyl chemical shift statistics extracted from the BioMagResBank. . 106

3.19 Extent of applied NOE distance and torsion restraints applied for minimalistic structure calculation protocol. 146

3.20 Deposited protein-ligand complexes employed for testing automated binding site mapping. 150

3.21 Description of adjustable parameters in ROSA. 152

3.22 Extent of available backbone chemical shifts for testing ROSA. 152

4.1 Backbone and sidechain chemical shift assignments for truncated (N-terminal) ICln. 156

4.2 Backbone chemical shifts for C-terminal ICln. 164

4.3 Measured $J_{C^{\gamma}C'}$ and $^{3}J_{C^{\gamma}N}$ couplings and associated uncertainties for truncated (N-terminal) ICln. 167

4.4 Applied conformational restraints and structural statistics for the 15 final models of ICln. 171

4.5 RMSD from the mean structure calculated from the final ensemble of 15 structures of ICln 172

4.6 Backbone chemical shift differences between *apo* ICln and the ICln/$\alpha_2\beta$ integrin peptide complex. 177

4.7 Backbone chemical shift differences observed in titration experiments with LSm4 as putative ICln ligand. 178

4.8 Backbone and sidechain chemical shift assignment of cyclophilin D. 182

4.9 Structural statistics for the final 15 Cyclophilin D structures. 192

4.10 RMS deviations to the mean structure of cyclophilin D. 193
4.11 Backbone and sidechain chemical shifts of *quail* LIM1(CRP2). 195
4.12 Structural statistics for the best final 15 LIM1/CRP2 structures. 201
4.13 RMS deviations to the mean structure of LIM1. 202
4.14 Backbone chemical shift assignment of brain abundant protein 1 (BASP1). . . . 205
4.15 Backbone chemical shift assignment of osteopontin (OPN). 211
4.16 Comparison of obtained automated and manual methyl chemical shift assignments.221
4.17 Quality of obtained structures using sparse input data. 226
4.18 Proteins tested and extent of input and deduced artificial restraints. 230
4.19 Quality of the obtained unconnected atomic 'cloud' protein images. 233
4.20 Extent of achieved automated sequential assignments. 240
4.21 Effects of limited precision of chemical shifts and completeness of chemical shift database. 247
4.22 Effects of structural errors in the atom density 'clouds' applied for sequential assignment. 249
4.23 Results of automated binding site mapping approach. 252

Publications

Papers

1. Schedlbauer A., Kontaxis G., Knig M., Frst J., Jakab M., Ritter M., Garavaglia L., Bott G., Meyer G., Paulmichl M. & Konrat R., 'Sequence-specific resonance assignments of ICln, an ion channel cloned from epithelial cells.' (2003) *J Biomol NMR*, 27, 399-400.

2. Fürst J., Schedlbauer A., Gandini R., Garavaglia M.L., Saino S., Gschwentner M., Sarg B., Lindner H., Jakab M., Ritter M., Bazzini C., Botta G., Meyer G., Kontaxis G., Tilly B.C., Konrat R. & Paulmichl, M. 'ICln159 folds into a pleckstrin homology domain-like structure. Interaction with kinases and the splicing factor LSm4' (2005) *J Biol Chem*, 280, 31276-31282.

3. Schedlbauer A., Hoffmann B., Kontaxis G., Rdisser S., Hommel U. & Konrat R. 'Automated backbone and side-chain assignment of mitochondrial matrix cyclophilin D' (2007) *J Biomol NMR*, 38, 267.

4. Schedlbauer A., Ozdowy P., Kontaxis G., Hartl M., Bister K. & Konrat R. (2008) 'Backbone assignment of osteopontin, a cytokine and cell attachment protein implicated in tumorigenesis', *Biomol. NMR Assign.*, 2(1), 29-31.

5. Schedlbauer A., Auer R., Ledolter K., Tollinger M., Kloiber K., Lichtenecker R., Ruedisser S., Hommel U., Schmid W., Konrat R. & Kontaxis G. (2008) 'Direct Methods and Residue Type Specific Isotope Labeling in NMR Structure Determination and Model-Driven Sequential Assignment' *J Biomol NMR*, 42, 111-127.

6. Schedlbauer A., Auer R., Kloiber K., Tollinger M. & Konrat R. (2008) "Auto-Correlation Analysis of NOESY Data provides Residue Compactness for Folded and Unfolded Proteins" *J Am. Chem. Soc.*, submitted.

7. Quintero R., Kloiber K., Schüler W., Schedlbauer A., Puffer B., Micura R., Kontaxis G., Konrat R. (2008) 'Automated Protein-Ligand Binding Site Mapping Using Unassigned Triple-Resonance Data', submitted.

8. Ozdowy P., Schedlbauer A., Bister K. & Konrat R. (2008) 'Molecular characterization of Basp1' in submission.

Poster Presentations

1. 'The solution structure of ICln a cytosolic protein that modulates volume regulated anion current associated with hypertonic cell swelling', *6^{th} Igler NMR Tage*, Congress center Igls, Austria, 24.02.2004-27.02.2004.

2. 'Rapid protein fold determination by NMR using proton density clouds and minimal NOE information' *7^{th} Igler NMR Tage*, University Center Obergurgl, Austria, 12.02.2006-15.02.2006.

3. 'Rapid protein fold determination by NMR using proton density clouds and minimal NOE information', *22^{th} International Conference on Magnetic Resonance in Biological Systems*, Goettingen, Germany, 20.08.2006-25.08.2006.

4. 'Model-Driven Protein Backbone Signal Assignment Using Proton Density Clouds and Minimal Triple Resonance Data', *Keystone Symposia on Structural Biology*, Snowbird Resort, Snowbird, Utah USA, 06.01.2007-11.01.2007

AWARDS

Poster Award for the contribution 'Rapid protein fold determination by NMR using proton density clouds and minimal NOE information' *7^{th} Igler NMR Tage*, University Center Obergurgl, Austria, 12.02.2006-15.02.2006.

Curriculum Vitae

Andreas Schedlbauer (Mag. rer. nat.)

CONTACT INFORMATION

Siegelgasse 21 Phone ++43-0650-5306463

A-2380 Perchtoldsdorf, Austria Email: andiatat@yahoo.de

PERSONAL DATA

Born: June 8^{th}, 1973 in vienna (austria)

Family status: single

Citizenship: austrian

Languages: german (mother tongue) and english

Personal interests: mountain hiking and biking, ski tours, swimming and soccer

EDUCATION

University

28.10.1993 Certificate of General Educational Development (University of Vienna)

01.03.1994 Begin studies of Biochemistry according degree programme A421 (University of Vienna)

30.03.2001 Diploma in Biochemistry

Theme*"NMR conformational analysis of the macrocycle Epothilon B and derivates"*

at the Department of organic Chemistry (Prof. Dr. Johannes Mulzer)

01.03.2002

to present Graduate research fellow with Prof. Dr. Robert Konrat

Department of Biomolecular Structural Chemistry, (University of Vienna)

Ph.D. Thesis *"Towards automated Structure determination"*

Development of methods for automated protein structure characterization

NMR structure analysis und dynamic of folded and partial unfolded biomolecules.

During period 01.04.2006 to 31.12.2007 this position was administrated by the Max Perutz Lab. LLC A-1030 Vienna. The defence of the Ph.D. Thesis will be accomplishised soon.

Schools

1987-1988 Technical school for structural design at the HTBLA ('technical colleg') A-2340 Moedling

1988-1992 Technical school for Biochemistry at the HTBLA for chemical Industry A-1170 Vienna

Courses

17.09.1995 Stacker lift truck licence at the institute for business development A-3100 St. Poelten

TEACHING

WS2001/02 Chemical lab class, (University of Vienna), course number 806943, 5 credits points.

EMPLOYMENT HISTORY

15.03.1993

to 15.04.1993 Immuno (now baxter) corporation (A-1210 Vienna) in the FSME vaccine production

01.06.1993

to 28.02.1994 Dragoco LLC (A-1230 Vienna) Quality control of flavor substances and essences

15.05.2001

to 28.02.2002 Department of Bioanorganic Chemistry at the group of Prof. Bernhard Keppler, University of Vienna as project collaborator with the theme *'Synthesis and spectroscopic characterization of tumor-repressive platine complexes'*

INTERNSHIPS

01.07.1989

to 31.07.1989 HOECHST corporation (A-1200 Wien) Production of washing agent additives

01.07.1990

to 31.07.1990 Stolllack corporation (A-2352 Guntramsdorf) in the paint development

01.07.1991

to 31.07.1991 Gebrueder Bablik (A-2345 Brunn/Gebirge) at the Galvanizing plant

01.08.1994

to 31.08.1994 Dragoco LLC (A-1230 Vienna) at the essence production

01.08.1995

to 31.08.1995 Dragoco LLC (A-1230 Vienna) Quality control of flavor substances and essences

01.08.1996

to 31.08.1996 Department of Organic Chemistry in the group Prof. Johann Mulzer University of Vienna *'Synthesis of chiral methyl groups'*

01.06.1998

to 15.04.1999 IKEA (A-2334 Voesendorf) function in store commission

MILITARY SERVICE

01.07.1992

to 28.02.1993 Corpsman at the austrian army

Bibliography

[1] T. Achsel, H. Brahms, B. Kastner, A. Bachi, M. Wilm, and R. Luehrmann (1999). 'A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro.' *EMBO J*, **18**(20), 5789–5802.

[2] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter (2003). Molekularbiologie der Zelle. Wiley-VCH.

[3] W. Arap, R. Pasqualini, and E. Ruoslahti (1998). 'Cancer treatment by targeted drug delivery to tumor vasculature in a mouse model.' *Science*, **279**(5349), 377–380.

[4] S. Ashkar, G. F. Weber, V. Panoutsakopoulou, M. E. Sanchirico, M. Jansson, S. Zawaideh, S. R. Rittling, D. T. Denhardt, M. J. Glimcher, and H. Cantor (2000). 'Eta-1 (osteopontin): an early component of type-1 (cell-mediated) immunity.' *Science*, **287**(5454), 860–864.

[5] R. A. Atkinson and V. Saudek (2002). 'The direct determination of protein structure by NMR without assignment.' *FEBS Lett*, **510**(1-2), 1–4.

[6] H. Barkhuijsen, R. de Beer, W. M. Bovee, J. H. Creyghton, and D. van Ormondt (1985). 'Application of linear prediction and singular value decomposition (LPSVD) to determine NMR frequencies and intensities from the FID.' *Magn Reson Med*, **2**(1), 86–89.

[7] A. Bax (2003). 'Weak alignment offers new NMR opportunities to study protein structure and dynamics.' *Protein Sci*, **12**(1), 1–16.

[8] K. J. Bayless, R. Salazar, and G. E. Davis (2000). 'RGD-dependent vacuolation and lumen formation observed during endothelial cell morphogenesis in three-dimensional fi brin matrices involves the alpha(v)beta(3) and alpha(5)beta(1) integrins.' *Am J Pathol*, **156**(5), 1673–1683.

[9] N. Blomberg, E. Baraldi, M. Nilges, and M. Saraste (1999). 'The PH superfold: a structural scaffold for multiple functions.' *Trends Biochem Sci*, **24**(11), 441–445.

[10] N. Blomberg, E. Baraldi, M. Sattler, M. Saraste, and M. Nilges (2000). 'Structure of a PH domain from the C. elegans muscle protein UNC-89 suggests a novel function.' *Structure*, **8**(10), 1079–1087.

[11] W. M. Bolstad (2004). Introduction to Bayesian Statistics. John Wiley New York.

[12] M. J. Bottomley, K. Salim, and G. Panayotou (1998). 'Phospholipid-binding protein domains.' *Biochim Biophys Acta*, **1436**(1-2), 165–183.

[13] L. Y. Bourguignon (2001). 'CD44-mediated oncogenic signaling and cytoskeleton activation during mammary tumor progression.' *J Mammary Gland Biol Neoplasia*, **6**(3), 287–297.

[14] D. Braaten, E. K. Franke, and J. Luban (1996). 'Cyclophilin A is required for the replication of group M human immunodefi ciency virus type 1 (HIV-1) and simian immunodefi ciency virus SIV(CPZ)GAB but not group O HIV-1 or other primate immunodefi ciency viruses.' *J Virol*, **70**(7), 4220–4227.

[15] A. L. Breeze (2000). 'Isotope-fi ltered NMR methods for the study of biomolecular structure and interactions.' *Prog. NMR Spectrosc.*, **36**, 323?372.

[16] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. S. S. Swaminathan, and M. Karplus (1983). 'CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations'. *J. Comp. Chem.*, **4**, 187–217.

[17] A. T. Brnger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, and G. L. Warren (1998). 'Crystallography & NMR system: A new software suite for macromolecular structure determination.' *Acta Crystallogr D Biol Crystallogr*, **54**(Pt 5), 905–921.

[18] A. G. Buckley (1976). 'Constrained Minimization Using Powell's Conjugacy Approach'. *SIAM Journal on Numerical Analysis*, **13**, 520–535.

[19] G. Buyse, T. Voets, J. Tytgat, C. D. Greef, G. Droogmans, B. Nilius, and J. Eggermont (1997). 'Expression of human pICln and ClC-6 in Xenopus oocytes induces an identical endogenous chloride conductance.' *J Biol Chem*, **272**(6), 3615–3621.

[20] B. Carpenter, K. J. Hill, M. Charalambous, K. J. Wagner, D. Lahiri, D. I. James, J. S. Andersen, V. Schumacher, B. Royer-Pokora, M. Mann, A. Ward, and S. G. E. Roberts (2004). 'BASP1 is a transcriptional cosuppressor for the Wilms' tumor suppressor protein WT1.' *Mol Cell Biol*, **24**(2), 537–549.

[21] D. A. Case (1998). 'The use of chemical shifts and their anisotropies in biomolecular structure determination.' *Curr Opin Struct Biol*, **8**(5), 624–630.

[22] D. S. Cassarino, R. H. Swerdlow, J. K. Parks, W. D. Parker, and J. P. Bennett (1998). 'Cyclosporin A increases resting mitochondrial membrane potential in SY5Y cells and reverses the depressed mitochondrial membrane potential of Alzheimer's disease cybrids.' *Biochem Biophys Res Commun*, **248**(1), 168–173.

[23] J. Cavanagh and M. Rance (1990). 'Sensitivity improvement in isotropic mixing (TOCSY) experiments'. *J. Mag. Res.*, **88**, 72–85.

[24] G. M. Clore, A. M. Gronenborn, M. Nilges, and C. A. Ryan (1987). 'Three-dimensional structure of potato carboxypeptidase inhibitor in solution. A study using nuclear magnetic resonance, distance geometry, and restrained molecular dynamics.' *Biochemistry*, **26**(24), 8012–8023.

[25] G. Cornilescu, F. Delaglio, and A. Bax (1999). 'Protein backbone angle restraints from searching a database for chemical shift and sequence homology.' *J Biomol NMR*, **13**(3), 289–302.

[26] M. Crompton, E. Barksby, N. Johnson, and M. Capano (2002). 'Mitochondrial intermembrane junctional complexes and their involvement in cell death.' *Biochimie*, **84**(2-3), 143–152.

[27] M. Crompton, S. Virji, V. Doyle, N. Johnson, and J. M. Ward (1999). 'The mitochondrial permeability transition pore.' *Biochem Soc Symp*, **66**, 167–179.

[28] I. B. Dawid, J. J. Breen, and R. Toyama (1998). 'LIM domains: multiple roles as adapters and functional modifi ers in protein interactions.' *Trends Genet*, **14**(4), 156–162.

[29] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax (1995). 'NMRPipe: a multidimensional spectral processing system based on UNIX pipes.' *J. Biomol. NMR*, **6**(3), 277–293.

[30] D. T. Denhardt, C. M. Giachelli, and S. R. Rittling (2001). 'Role of osteopontin in cellular signaling and toxicant injury.' *Annu Rev Pharmacol Toxicol*, **41**, 723–749.

[31] H. Diao, S. Kon, K. Iwabuchi, C. Kimura, J. Morimoto, D. Ito, T. Segawa, M. Maeda, J. Hamuro, T. Nakayama, M. Taniguchi, H. Yagita, L. V. Kaer, K. One, D. Denhardt, S. Rittling, and T. Uede (2004). 'Osteopontin as a mediator of NKT cell function in T cell-mediated liver diseases.' *Immunity*, **21**(4), 539–550.

[32] T. Diercks, M. Daniels, and R. Kaptein (2005). 'Extended flip-back schemes for sensitivity enhancement in multidimensional HSQC-type out-and-back experiments.' *J Biomol NMR*, **33**(4), 243–259.

[33] P. Dosset, J. C. Hus, M. Blackledge, and D. Marion (2000). 'Effi cient analysis of macromolecular rotational diffusion from heteronuclear relaxation data.' *J Biomol NMR*, **16**(1), 23–28.

[34] B. M. Duggan, G. B. Legge, H. J. Dyson, and P. E. Wright (2001). 'SANE (Structure Assisted NOE Evaluation): an automated model-based approach for NOE assignment.' *J Biomol NMR*, **19**(4), 321–329.

[35] P. Duhamel and M. Vetterli (1990). 'Fast fourier transforms: A tutorial review and a state of the art'. *Signal Processing*, **19**, 259–299.

[36] A. Ebel, W. Dreher, and D. Leibfritz (2006). 'Effects of zero-fi lling and apodization on spectral integrals in discrete Fourier-transform spectroscopy of noisy data.' *J Magn Reson*, **182**(2), 330–338.

[37] C. Eichmueller, M. Tollinger, B. Kraeutler, and R. Konrat (2001). 'Mapping the ligand binding site at protein side-chains in protein-ligand complexes through NOE difference spectroscopy.' *J Biomol NMR*, **20**(3), 195–202.

[38] F. Emma, R. Sanchez-Olea, and K. Strange (1998). 'Characterization of pI(Cln) binding proteins: identifi cation of p17 and assessment of the role of acidic domains in mediating protein-protein interactions.' *Biochim Biophys Acta*, **1404**(3), 321–328.

[39] N. A. Farrow, R. Muhandiram, A. U. Singer, S. M. Pascal, C. M. Kay, G. Gish, S. E. Shoelson, T. Pawson, J. D. Forman-Kay, and L. E. Kay (1994). 'Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation.' *Biochemistry*, **33**(19), 5984–6003.

[40] L. W. Fisher, D. A. Torchia, B. Fohr, M. F. Young, and N. S. Fedarko (2001). 'Flexible structures of SIBLING proteins, bone sialoprotein, and osteopontin.' *Biochem Biophys Res Commun*, **280**(2), 460–465.

[41] C. Fletcher, D. Jones, R. Diamond, and D. Neuhaus (1996). 'Treatment of NOE constraints involving equivalent or nonstereoassigned protons in calculations of biomacromolecular structures.' *J. Biomol. NMR*, **8**, 292?310.

[42] W. J. Friesen, S. Paushkin, A. Wyce, S. Massenet, G. S. Pesiridis, G. V. Duyne, J. Rappsilber, M. Mann, and G. Dreyfuss (2001). 'The methylosome, a 20S complex containing JBP1 and pICln, produces dimethylarginine-modifi ed Sm proteins.' *Mol Cell Biol*, **21**(24), 8289–8300.

[43] J. Fuerst, C. Bazzini, M. Jakab, G. Meyer, M. Knig, M. Gschwentner, M. Ritter, A. Schmarda, G. Bott, R. Benz, P. Deetjen, and M. Paulmichl (2000). 'Functional reconstitution of ICln in lipid bilayers.' *Pflugers Arch*, **440**(1), 100–115.

[44] J. Fuerst, M. Jakab, M. Knig, M. Ritter, M. Gschwentner, J. Rudzki, J. Danzl, M. Mayer, C. M. Burtscher, J. Schirmer, B. Maier, M. Nairz, S. Chwatal, and M. Paulmichl (2000). 'Structure and function of the ion channel ICln.' *Cell Physiol Biochem*, **10**(5-6), 329–334.

[45] J. Frst, A. Schedlbauer, R. Gandini, M. L. Garavaglia, S. Saino, M. Gschwentner, B. Sarg, H. Lindner, M. Jakab, M. Ritter, C. Bazzini, G. Botta, G. Meyer, G. Kontaxis, B. C. Tilly, R. Konrat, and M. Paulmichl (2005). 'ICln159 folds into a pleckstrin homology domain-like structure. Interaction with kinases and the splicing factor LSm4.' *J Biol Chem*, **280**(35), 31276–31282.

[46] M. L. Garavaglia, S. Rodighiero, C. Bertocchi, R. Manfredi, J. Fuerst, M. Gschwentner, M. Ritter, C. Bazzini, G. Bott, M. Jakab, G. Meyer, and M. Paulmichl (2002). 'ICln channels reconstituted in heart-lipid bilayer are selective to chloride.' *Pflugers Arch*, **443**(5-6), 748–753.

[47] A. Gericke, C. Qin, L. Spevak, Y. Fujimoto, W. T. Butler, E. S. Srensen, and A. L. Boskey (2005). 'Importance of phosphorylation for osteopontin regulation of biomineralization.' *Calcif Tissue Int*, **77**(1), 45–54.

[48] T. D. Goddard and D. G. Kneller (2002). 'SPARKY version 3.114'. Technical report, University of California.

[49] P. Gouet, E. Courcelle, D. I. Stuart, and F. Metoz (1999). 'ESPript: analysis of multiple sequence alignments in PostScript.' *Bioinformatics*, **15**(4), 305–308.

[50] A. Grishaev and M. Llins (2002). 'CLOUDS, a protocol for deriving a molecular proton density via NMR.' *Proc. Natl. Acad. Sci. U S A*, **99**(10), 6707–6712.

[51] S. Grzesiek, J. Anglister, and A. Bax (1992). 'Correlation of backbone Amide and Aliphatic Side-Chain Resonances in13C/15N-Enrieched Proteins by Isotropic Mixing of 13C Magnetization'. *J. Magn. Reson. B*, **101**, 114–119.

[52] S. Grzesiek and A. Bax (1992). 'Correlation of Backbone Amide and Aliphatic Side-Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance NMR'. *J. Am. Chem. Soc.*, **114**, 6291–6293.

[53] S. Grzesiek, G. W. Vuister, and A. Bax (1993). 'A simple and sensitive experiment for measurement of JCC couplings between backbone carbonyl and methyl carbons in isotopically enriched proteins.' *J Biomol NMR*, **3**(4), 487–493.

[54] M. Gschwentner, U. O. Nagl, E. Wll, A. Schmarda, M. Ritter, and M. Paulmichl (1995). 'Antisense oligonucleotides suppress cell-volume-induced activation of chloride channels.' *Pflugers Arch*, **430**(4), 464–470.

[55] M. Gschwentner, A. Susanna, A. Schmarda, A. Laich, U. O. Nagl, H. Ellemunter, P. Deetjen, J. Frick, and M. Paulmichl (1996). 'ICln: a chloride channel paramount for cell volume regulation.' *J Allergy Clin Immunol*, **98**(5 Pt 2), S98–101; discussion S105–6.

[56] P. Guentert, K. Berndt, and K. Wuethrich (1993). 'The program ASNO for computer-supported collection of NOE upper distance constraints as input for protein structure determination'. *J. Biomol. NMR*, **3**, 601–606.

[57] P. Guentert, W. Braun, and K. Wuethrich (1991). 'Effi cient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA.' *J Mol Biol*, **217**(3), 517–530.

[58] P. Guentert, C. Mumenthaler, and K. Wuethrich (1997). 'Torsion angle dynamics for NMR structure calculation with the new program DYANA.' *J Mol Biol*, **273**(1), 283–298.

[59] M. Hartl, T. Matt, W. Schler, G. Siemeister, G. Kontaxis, K. Kloiber, R. Konrat, and K. Bister (2003). 'Cell transformation by the v-myc oncogene abrogates c-Myc/Max-mediated suppression of a C/EBP beta-dependent lipocalin gene.' *J Mol Biol*, **333**(1), 33–46.

[60] T. F. Havel, I. D. Kuntz, and G. M. Crippen (1983). 'The combinatorial distance geometry method for the calculation of molecular conformation. I. A new approach to an old problem.' *J Theor Biol*, **104**(3), 359–381.

[61] T. Hayano, N. Takahashi, S. Kato, N. Maki, and M. Suzuki (1991). 'Two distinct forms of peptidylprolyl-cis-trans-isomerase are expressed separately in periplasmic and cytoplasmic compartments of Escherichia coli cells.' *Biochemistry*, **30**(12), 3041–3048.

[62] T. Herrmann, P. Guentert, and K. Wuethrich (2002). 'Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA.' *J Mol Biol*, **319**(1), 209–227.

[63] T. K. Hitchens, J. A. Lukin, Y. Zhan, S. A. McCallum, and G. S. Rule (2003). 'MONTE: An automated Monte Carlo based approach to nuclear magnetic resonance assignment of proteins.' *J. Biomol. NMR*, **25**(1), 1–9.

[64] J. Hoch and A. S. Stern (1996). NMR Data Processing. Wiley-Liss.

[65] B. Hoffmann, C. Eichmller, O. Steinhauser, and R. Konrat (2005). 'Rapid Assessment of Protein Structural Stability and Fold Validation via NMR'. *Methods in Enzymology*, **394**, 3–631.

[66] T. J. Hubbard, A. G. Murzin, S. E. Brenner, and C. Chothia (1997). 'SCOP: a structural classifi cation of proteins database.' *Nucleic Acids Res*, **25**(1), 236–239.

[67] M. D. Hubert, I. Levitan, M. M. Hoffman, M. Zraggen, M. E. Hofreiter, and S. S. Garber (2000). 'Modulation of volume regulated anion current by I(Cln).' *Biochim Biophys Acta*, **1466**(1-2), 105–114.

[68] P. E. Hughes, T. E. O'Toole, J. Ylnne, S. J. Shattil, and M. H. Ginsberg (1995). 'The conserved membrane-proximal region of an integrin cytoplasmic domain specifi es ligand binding affi nity.' *J Biol Chem*, **270**(21), 12411–12417.

[69] J. Iizuka, Y. Katagiri, N. Tada, M. Murakami, T. Ikeda, M. Sato, K. Hirokawa, S. Okada, M. Hatano, T. Tokuhisa, and T. Uede (1998). 'Introduction of an osteopontin gene confers the increase in B1 cell population and the production of anti-DNA autoantibodies.' *Lab Invest*, **78**(12), 1523–1533.

[70] T. Ikeya, T. Terauchi, P. Gntert, and M. Kainosho (2006). 'Evaluation of stereo-array isotope labeling (SAIL) patterns for automated structural analysis of proteins with CYANA.' *Magn Reson Chem*, **44 Spec No**, S152–S157.

[71] B. A. Johnson (2004). 'Using NMRView to visualize and analyze the NMR spectra of macromolecules.' *Methods Mol. Biol.*, **278**, 313–352.

[72] Y.-S. Jung and M. Zweckstetter (2004). 'Mars – robust automatic backbone assignment of proteins.' *J Biomol NMR*, **30**(1), 11–23.

[73] J. L. Kadrmas and M. C. Beckerle (2004). 'The LIM domain: from the cytoskeleton to the nucleus.' *Nat Rev Mol Cell Biol*, **5**(11), 920–931.

[74] C. Kambach, S. Walke, R. Young, J. M. Avis, E. de la Fortelle, V. A. Raker, R. Lhrmann, J. Li, and K. Nagai (1999). 'Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs.' *Cell*, **96**(3), 375–387.

[75] L. Kay, M. Ikura, R. Tschudin, and A. Bax. (1990). 'The HNCA experiment'. *J. Magn. Reson.*, **89**, 496–514.

[76] L. Kay, L. Nicholson, F. Delaglio, A. Bax, and D. Torchia (1992). 'PULSE SEQUENCES FOR REMOVAL OF THE EFFECTS OF CROSS-CORRELATION BETWEEN DIPOLAR AND CHEMICAL-SHIFT ANISOTROPY RELAXATION MECHANISM ON THE MEASUREMENT OF HETERONUCLEAR T1 AND T2 VALUES IN PROTEINS'. *J. Mag. Resn.*, **97**, 357.

[77] L. E. Kay (2005). 'NMR studies of protein structure and dynamics.' *J Magn Reson*, **173**(2), 193–207.

[78] L. E. Kay and K. H. Gardner (1997). 'Solution NMR spectroscopy beyond 25 kDa.' *Curr Opin Struct Biol*, **7**(5), 722–731.

[79] L. E. Kay, P. Keifer, and T. Saarinen. (1992). 'Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity'. *J. Am. Chem. Soc.*, **114**, 10663–10665.

[80] C. C. Kazanecki, D. J. Uzwiak, and D. T. Denhardt (2007). 'Control of osteopontin signaling and function by post-translational phosphorylation and protein folding.' *J Cell Biochem*, **102**(4), 912–924.

[81] J. O. Kim, M. M. Nau, K. A. Allikian, T. P. Mkel, K. Alitalo, B. E. Johnson, and M. J. Kelley (1998). 'Co-amplifi cation of a novel cyclophilin-like gene (PPIE) with L-myc in small cell lung cancer cell lines.' *Oncogene*, **17**(8), 1019–1026.

[82] A. J. Koletsky, M. W. Harding, and R. E. Handschumacher (1986). 'Cyclophilin: distribution and variant properties in normal and neoplastic tissues.' *J Immunol*, **137**(3), 1054–1059.

[83] R. Konrat, R. Weiskirchen, B. Krutler, and K. Bister (1997). 'Solution structure of the carboxyl-terminal LIM domain from quail cysteine-rich protein CRP2.' *J Biol Chem*, **272**(18), 12001–12007.

[84] G. Kontaxis, R. Konrat, B. Krutler, R. Weiskirchen, and K. Bister (1998). 'Structure and intramodular dynamics of the amino-terminal LIM domain from quail cysteine- and glycine-rich protein CRP2.' *Biochemistry*, **37**(20), 7127–7134.

[85] R. Koradi, M. Billeter, and K. Wuethrich (1996). 'MOLMOL: a program for display and analysis of macromolecular structures.' *J Mol Graph*, **14**(1), 51–5, 29–32.

[86] R. Koradi, M. Billeter, and K. Wuthrich (1999). 'MOLMOL: a program for display and analysis of macromolecular structures.' *J. Mol. Graphics*, **14**, 51–55.

[87] J. L. Kosa, J. W. Michelsen, H. A. Louis, J. I. Olsen, D. R. Davis, M. C. Beckerle, and D. R. Winge (1994). 'Common metal ion coordination in LIM domain proteins.' *Biochemistry*, **33**(2), 468–477.

[88] G. B. Krapivinsky, M. J. Ackerman, E. A. Gordon, L. D. Krapivinsky, and D. E. Clapham (1994). 'Molecular characterization of a swelling-induced chloride conductance regulatory protein, pICln.' *Cell*, **76**(3), 439–448.

[89] P. J. Kraulis (1994). 'Protein three-dimensional structure determination and sequence-specifi c assignment of 13C and 15N-separated NOE data. A novel real-space ab initio approach.' *J Mol Biol*, **243**(4), 696–718.

[90] P. Kumar, P. J. Mark, B. K. Ward, R. F. Minchin, and T. Ratajczak (2001). 'Estradiol-regulated expression of the immunophilins cyclophilin 40 and FKBP52 in MCF-7 breast cancer cells.' *Biochem Biophys Res Commun*, **284**(1), 219–225.

[91] J. Kuszewski, A. M. Gronenborn, and G. M. Clore (1996). 'Improving the quality of NMR and crystallographic protein structures by means of a conformational database potential derived from structure databases.' *Protein Sci*, **5**(6), 1067–1080.

[92] J. Kuszewski, A. M. Gronenborn, and G. M. Clore (1999). 'Improving the Packing and Accuracy of NMR Structures with a Pseudopotential for the Radius of Gyration'. *J. Am. Chem. Soc.*, **121**, 2337–2338.

[93] J. Kyte and R. F. Doolittle (1982). 'A simple method for displaying the hydropathic character of a protein.' *J Mol Biol*, **157**(1), 105–132.

[94] D. Larkin, D. Murphy, D. F. Reilly, M. Cahill, E. Sattler, P. Harriott, D. J. Cahill, and N. Moran (2004). 'ICln, a novel integrin alphaIIbbeta3-associated protein, functionally regulates platelet activation.' *J Biol Chem*, **279**(26), 27286–27293.

[95] R. A. Laskowski, D. S. Moss, and J. M. Thornton (1993). 'Main-chain bond lengths and bond angles in protein structures.' *J Mol Biol*, **231**(4), 1049–1067.

[96] R. A. Laskowski, J. A. Rullmannn, M. W. MacArthur, R. Kaptein, and J. M. Thornton (1996). 'AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR.' *J Biomol NMR*, **8**(4), 477–486.

[97] T. Lassmann and E. L. L. Sonnhammer (2005). 'Kalign–an accurate and fast multiple sequence alignment algorithm.' *BMC Bioinformatics*, **6**, 298.

[98] D. M. LeMaster (1988). 'Protein NMR resonance assignment by isotropic mixing experiments on random fractionally deuterated samples.' *FEBS Lett*, **233**(2), 326–330.

[99] M. A. Lemmon, K. M. Ferguson, and C. S. Abrams (2002). 'Pleckstrin homology domains and the cytoskeleton.' *FEBS Lett*, **513**(1), 71–76.

[100] C. Li, S. Breton, R. Morrison, C. L. Cannon, F. Emma, R. Sanchez-Olea, C. Bear, and K. Strange (1998). 'Recombinant pICln forms highly cation-selective channels when reconstituted into artifi - cial and biological membranes.' *J Gen Physiol*, **112**(6), 727–736.

[101] R. Lichtenecker, M. L. Ludwiczek, W. Schmid, and R. Konrat (2004). 'Simplifi cation of protein NOESY spectra using bioorganic precursor synthesis and NMR spectral editing.' *J Am Chem Soc*, **126**(17), 5348–5349.

[102] J. P. Linge, M. Habeck, W. Rieping, and M. Nilges (2003). 'ARIA: automated NOE assignment and NMR structure calculation.' *Bioinformatics*, **19**(2), 315–316.

[103] G. Lipari and A. Szabo (1982). 'Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity'. *J . Am. Chem. Soc.*, **104**, 4546 – 4559.

[104] G. Lipari and A. Szabo (1982). 'Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results'. *J . Am. Chem. Soc.*, **104**, 4559 – 4570.

[105] Q. Liu, U. Fischer, F. Wang, and G. Dreyfuss (1997). 'The spinal muscular atrophy disease gene product, SMN, and its associated protein SIP1 are in a complex with spliceosomal snRNP proteins.' *Cell*, **90**(6), 1013–1021.

[106] J. A. Lukin, A. P. Gove, S. N. Talukdar, and C. Ho (1997). 'Automated probabilistic method for assigning backbone resonances of (13C,15N)-labeled proteins.' *J Biomol NMR*, **9**(2), 151–166.

[107] M. W. MacArthur and J. M.Thornton (1993). 'Conformational analysis of protein structures derived from NMR data'. *Proteins*, **17**, 232–251.

[108] D. Malmodin, C. H. M. Papavoine, and M. Billeter (2003). 'Fully automated sequence-specifi c resonance assignments of hetero- nuclear protein spectra.' *J Biomol NMR*, **27**(1), 69–79.

[109] D. Marion, M. Ikura, and R. T. andd A. Bax (1989). 'Rapid recording of 2D NMR spectra without phase cycling. Application to the study of hydrogen exchange in proteins'. *J. Magn. Reson*, **85**, 393–399.

[110] S.-J. Marrink and H. J. C. Berendsen (1994). 'Simulation of Water Transport through a Lipid Membrane'. *J. Phys. Chem.*, **98**, 4155–4168.

[111] I. Marzo, C. Brenner, N. Zamzami, S. A. Susin, G. Beutner, D. Brdiczka, R. Rmy, Z. H. Xie, J. C. Reed, and G. Kroemer (1998). 'The permeability transition pore complex: a target for apoptosis regulation by caspases and bcl-2-related proteins.' *J Exp Med*, **187**(8), 1261–1271.

[112] A. E. Mayes, L. Verdone, P. Legrain, and J. D. Beggs (1999). 'Characterization of Sm-like proteins in yeast and their association with U6 snRNA.' *EMBO J*, **18**(15), 4321–4331.

[113] S. A. McCallum, T. K. Hitchens, and G. S. Rule (1999). 'Solution structure of the carboxyl terminus of a human class Mu glutathione S-transferase: NMR assignment strategies in large proteins.' *J Mol Biol*, **285**(5), 2119–2132.

[114] L. McIntosh and F. Dahlquist (1990). 'Biosynthetic incorporation of 15N and 13C for assignment and interpretation of nuclear magnetic resonance spectra of proteins.' *Q. Rev. Biophys.*, **23**, 1?38.

[115] M. D. McKee and A. Nanci (1995). 'Osteopontin and the bone remodeling sequence. Colloidal-gold immunocytochemistry of an interfacial extracellular matrix protein.' *Ann N Y Acad Sci*, **760**, 177–189.

[116] G. Meister, C. Eggert, D. Buehler, H. Brahms, C. Kambach, and U. Fischer (2001). 'Methylation of Sm proteins by a complex containing PRMT5 and the putative U snRNP assembly factor pICln.' *Curr Biol*, **11**(24), 1990–1994.

[117] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). 'Equation of state calculation by fast computing machines'. *Journal of Chemical Physics*, **21**, 1087?1092.

[118] G. Meyer, S. Rodighiero, F. Guizzardi, C. Bazzini, G. Bott, C. Bertocchi, L. Garavaglia, S. Dossena, R. Manfredi, C. Sironi, A. Catania, and M. Paulmichl (2004). 'Volume-regulated Cl- channels in human pleural mesothelioma cells.' *FEBS Lett*, **559**(1-3), 45–50.

[119] J. W. Michelsen, K. L. Schmeichel, M. C. Beckerle, and D. R. Winge (1993). 'The LIM motif defi nes a specifi c zinc-binding protein domain.' *Proc Natl Acad Sci U S A*, **90**(10), 4404–4408.

[120] R. T. Miller, S. P. Anderson, J. C. Corton, and R. C. Cattley (2000). 'Apoptosis, mitosis and cyclophilin-40 expression in regressing peroxisome proliferator-induced adenomas.' *Carcinogenesis*, **21**(4), 647–652.

[121] J. W. Montague, F. M. Hughes, and J. A. Cidlowski (1997). 'Native recombinant cyclophilins A, B, and C degrade DNA independently of peptidylprolyl cis-trans-isomerase activity. Potential roles of cyclophilins in apoptosis.' *J Biol Chem*, **272**(10), 6677–6684.

[122] G. A. Morris and R. Freeman (1979). 'Insensitive Nuclei Enhanced by Polarization Transfer'. *J. Am. Chem. Soc.*, **101**, 760–762.

[123] M. I. Mosevitsky, J. P. Capony, S. GYu, V. A. Novitskaya, P. AYu, and V. V. Zakharov (1997). 'The BASP1 family of myristoylated proteins abundant in axonal termini. Primary structure analysis and physico-chemical properties.' *Biochimie*, **79**(6), 373–384.

[124] D. R. Muhandiram and L. E. Kay (1994). 'Gradient-Enhanced Triple-Resonance Three-Dimensional NMR Experiments with Improved Sensitivity'. *J. Magn. Res. B*, **103**, 203–216.

[125] C. Mumenthaler, P. Gntert, W. Braun, and K. Wthrich (1997). 'Automated combined assignment of NOESY spectra and three-dimensional protein structure determination.' *J Biomol NMR*, **10**(4), 351–362.

[126] D. Neuhaus and M. Williamson (2000). The Nuclear Overhauser Effect in Structural and Conformational Analysis. John Wiley & Sons.

[127] M. Nilges, A. M. Gronenborn, A. T. Brnger, and G. M. Clore (1988). 'Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints. Application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2.' *Protein Eng*, **2**(1), 27–38.

[128] M. Nilges, M. J. Macias, S. I. O'Donoghue, and H. Oschkinat (1997). 'Automated NOESY interpretation with ambiguous distance restraints: the refi ned NMR solution structure of the pleckstrin homology domain from beta-spectrin.' *J Mol Biol*, **269**(3), 408–422.

[129] B. Nilius, J. Eggermont, T. Voets, and G. Droogmans (1996). 'Volume-activated Cl- channels.' *Gen Pharmacol*, **27**(7), 1131–1140.

[130] Y. Okada (1997). 'Volume expansion-sensing outward-rectifi er Cl- channel: fresh start to the molecular identity and volume sensor.' *Am J Physiol*, **273**(3 Pt 1), C755–C789.

[131] S. O'Neill, A. Robinson, A. Deering, M. Ryan, D. J. Fitzgerald, and N. Moran (2000). 'The platelet integrin alpha IIbbeta 3 has an endogenous thiol isomerase activity.' *J Biol Chem*, **275**(47), 36984–36990.

[132] S. J. Opella, A. Nevzorov, M. F. Mesleb, and F. M. Marassi (2002). 'Structure determination of membrane proteins by NMR spectroscopy.' *Biochem Cell Biol*, **80**(5), 597–604.

[133] M. Ottiger, O. Zerbe, P. Gntert, and K. Wthrich (1997). 'The NMR solution conformation of unligated human cyclophilin A.' *J Mol Biol*, **272**(1), 64–81.

[134] G. Otting, R. Grtter, W. Leupin, C. Minganti, K. N. Ganesh, B. S. Sproat, M. J. Gait, and K. Wthrich (1987). 'Sequential NMR assignments of labile protons in DNA using two-dimensional nuclear-Overhauser-enhancement spectroscopy with three jump-and-return pulse sequences.' *Eur J Biochem*, **166**(1), 215–220.

[135] H. D. Ou, H. C. Lai, Z. Serber, and V. Dtsch (2001). 'Effi cient identifi cation of amino acid types for fast protein backbone assignments.' *J Biomol NMR*, **21**(3), 269–273.

[136] S. C. Panchal, N. S. Bhavesh, and R. V. Hosur (2001). 'Improved 3D triple resonance experiments, HNN and HN(C)N, for HN and 15N sequential correlations in (13C, 15N) labeled proteins: application to unfolded proteins.' *J Biomol NMR*, **20**(2), 135–147.

[137] M. Paulmichl, Y.Li, K. Wickmann, M. Ackerman, E. Peralta, and D. Clapham (1992). 'New mammalian chloride channel identifi ed by expression cloning.' *Nature*, **356**, 238–241.

[138] K. Pervushin (2000). 'Impact of transverse relaxation optimized spectroscopy (TROSY) on NMR as a technique in structural biology.' *Q Rev Biophys*, **33**(2), 161–197.

[139] M. Piotto, V. Saudek, and V. Sklenr (1992). 'Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions.' *J Biomol NMR*, **2**(6), 661–665.

[140] W. S. Price, F. Elwinger, C. Vigouroux, and P. Stilbs (2002). 'PGSE-WATERGATE, a new tool for NMR diffusion-based studies of ligand-macromolecule binding'. *Magnetic Resonance in Chemistry*, **40**, 391 – 395.

[141] W. T. Pu, G. B. Krapivinsky, L. Krapivinsky, and D. E. Clapham (1999). 'pICln inhibits snRNP biogenesis by binding core spliceosomal proteins.' *Mol Cell Biol*, **19**(6), 4113–4120.

[142] W. T. Pu, K. Wickman, and D. E. Clapham (2000). 'ICln is essential for cellular and early embryonic viability.' *J Biol Chem*, **275**(17), 12363–12366.

[143] C. Prez, F. Lhr, H. Rterjans, and J. M. Schmidt (2001). 'Self-consistent Karplus parametrization of 3J couplings depending on the polypeptide side-chain torsion chi1.' *J Am Chem Soc*, **123**(29), 7081–7093.

[144] D. C. Radisky, M. C. Babcock, and J. Kaplan (1999). 'The yeast frataxin homologue mediates mitochondrial iron efflux. Evidence for a mitochondrial iron cycle.' *J Biol Chem*, **274**(8), 4497–4499.

[145] S. Rajesh, D. Nietlispach, H. Nakayama, K. Takio, E. D. Laue, T. Shibata, and Y. Ito (2003). 'A novel method for the biosynthesis of deuterated proteins with selective protonation at the aromatic rings of Phe, Tyr and Trp.' *J Biomol NMR*, **27**(1), 81–86.

[146] H. Rangaswami, A. Bulbule, and G. C. Kundu (2006). 'Nuclear factor inducing kinase: a key regulator in osteopontin- induced MAPK/IkappaB kinase dependent NF-kappaB-mediated promatrix metalloproteinase-9 activation.' *Glycoconj J*, **23**(3-4), 221–232.

[147] H. Rangaswami, A. Bulbule, and G. C. Kundu (2006). 'Osteopontin: role in cell signaling and cancer progression.' *Trends Cell Biol*, **16**(2), 79–87.

[148] C. Redfi eld, L. J. Smith, J. Boyd, G. M. Lawrence, R. G. Edwards, R. A. Smith, and C. M. Dobson (1991). 'Secondary structure and topology of human interleukin 4 in solution.' *Biochemistry*, **30**(46), 11029–11035.

[149] J. L. Risler, M. O. Delorme, H. Delacroix, and A. Henaut (1988). 'Amino acid substitutions in structurally related proteins. A pattern recognition approach. Determination of a new and effi cient scoring matrix.' *J Mol Biol*, **204**(4), 1019–1029.

[150] L. R. Rodrigues, J. A. Teixeira, F. L. Schmitt, M. Paulsson, and H. Lindmark-Mnsson (2007). 'The role of osteopontin in tumor progression and metastasis in breast cancer.' *Cancer Epidemiol Biomarkers Prev*, **16**(6), 1087–1097.

[151] S. Rogers, R. Wells, and M. Rechsteiner (1986). 'Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis.' *Science*, **234**(4774), 364–368.

[152] M. K. Rosen, K. H. Gardner, R. C. Willis, W. E. Parris, T. Pawson, and L. E. Kay (1996). 'Selective methyl group protonation of perdeuterated proteins.' *J Mol Biol*, **263**(5), 627–636.

[153] D. Roy, D. Ghosh, and S. Gupta-Bhattacharya (2003). 'Homology modeling of allergenic cyclophilins: IgE-binding site and structural basis of cross-reactivity.' *Biochem Biophys Res Commun*, **307**(2), 422–429.

[154] S. Rucker and A. Shaka (1989). 'Broadband homonuclear cross polarization in 2D N.M.R. using DIPSI-2.' *Mol. Phys.*, **68**, 509?517.

[155] B. Ryffel, G. Woerly, B. Greiner, B. Haendler, M. J. Mihatsch, and B. M. Foxwell (1991). 'Distribution of the cyclosporine binding protein cyclophilin in human tissues.' *Immunology*, **72**(3), 399–404.

[156] R. Sanchez-Olea, F. Emma, M. Coghlan, and K. Strange (1998). 'Characterization of pICln phosphorylation state and a pICln-associated protein kinase.' *Biochim Biophys Acta*, **1381**(1), 49–60.

[157] A. Sardini, J. S. Amey, K. H. Weylandt, M. Nobles, M. A. Valverde, and C. F. Higgins (2003). 'Cell volume regulation and swelling-activated chloride channels.' *Biochim Biophys Acta*, **1618**(2), 153–162.

[158] A. Schedlbauer, B. Hoffmann, G. Kontaxis, S. Rdisser, U. Hommel, and R. Konrat (2007). 'Automated backbone and side-chain assignment of mitochondrial matrix cyclophilin D.' *J Biomol NMR*, **38**(3), 267.

[159] A. Schedlbauer, G. Kontaxis, M. Knig, J. Frst, M. Jakab, M. Ritter, L. Garavaglia, G. Bott, G. Meyer, M. Paulmichl, and R. Konrat (2003). 'Sequence-specifi c resonance assignments of ICln, an ion channel cloned from epithelial cells.' *J Biomol NMR*, **27**(4), 399–400.

[160] D. Schlatter, R. Thoma, E. Kueng, M. Sthile, F. Mueller, E. Boroni, and M. Hennig (2005). 'Crystal engineering yields crystals of cyclophilin D diffracting to 1.7 resolution'. *Acta. Crystallogr. Sect.D*, **61**, 513.

[161] S. L. Schreiber (1992). 'Immunophilin-sensitive protein phosphatase action in cell signaling pathways.' *Cell*, **70**(3), 365–368.

[162] M. Schubert, H. Oschkinat, and P. Schmieder (2001). 'MUSIC, selective pulses, and tuned delays: amino acid type-selective (1)H-(15)N correlations, II.' *J Magn Reson*, **148**(1), 61–72.

[163] S. Schwarzinger, G. J. Kroon, T. R. Foss, J. Chung, P. E. Wright, and H. J. Dyson (2001). 'Sequence-dependent correction of random coil NMR chemical shifts.' *J Am Chem Soc*, **123**(13), 2970–2978.

[164] C. D. Schwieters, J. J. Kuszewski, N. Tjandra, and G. M. Clore (2003). 'The Xplor-NIH NMR molecular structure determination package.' *J. Magn. Reson.*, **160**(1), 65–73.

[165] M. J. Sippl (1990). 'Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins.' *J Mol Biol*, **213**(4), 859–883.

[166] M. J. Sippl (1993). 'Boltzmann's principle, knowledge-based mean fi elds and protein folding. An approach to the computational determination of protein structures.' *J Comput Aided Mol Des*, **7**(4), 473–501.

[167] M. J. Sippl (1993). 'Recognition of errors in three-dimensional structures of proteins.' *Proteins*, **17**(4), 355–362.

[168] M. J. Sippl (1995). 'Knowledge-based potentials for proteins.' *Curr Opin Struct Biol*, **5**(2), 229–235.

[169] V. Sklenr, B. R. Brooks, G. Zon, and A. Bax (1987). 'Absorption mode two-dimensional NOE spectroscopy of exchangeable protons in oligonucleotides.' *FEBS Lett*, **216**(2), 249–252.

[170] M. Y. Speer, Y.-C. Chien, M. Quan, H.-Y. Yang, H. Vali, M. D. McKee, and C. M. Giachelli (2005). 'Smooth muscle cells defi cient in osteopontin have enhanced susceptibility to calcifi cation in vitro.' *Cardiovasc Res*, **66**(2), 324–333.

[171] S. Spera, M. Ikura, and A. Bax (1991). 'Measurement of the exchange rates of rapidly exchanging amide protons: application to the study of calmodulin and its complex with a myosin light chain kinase fragment.' *J Biomol NMR*, **1**(2), 155–165.

[172] J. P. Steiner, G. S. Hamilton, D. T. Ross, H. L. Valentine, H. Guo, M. A. Connolly, S. Liang, C. Ramsey, J. H. Li, W. Huang, P. Howorth, R. Soni, M. Fuller, H. Sauer, A. C. Nowotnik, and P. D. Suzdak (1997). 'Neurotrophic immunophilin ligands stimulate structural and functional recovery in neurodegenerative animal models.' *Proc Natl Acad Sci U S A*, **94**(5), 2019–2024.

[173] T. O. Street, D. W. Bolen, and G. D. Rose (2006). 'A molecular mechanism for osmolyte-induced protein stability.' *Proc Natl Acad Sci U S A*, **103**(38), 13997–14002.

[174] I. Snchez-Garca and T. H. Rabbitts (1994). 'The LIM domain: a new structural motif found in zinc-fi nger-like proteins.' *Trends Genet*, **10**(9), 315–320.

[175] A. Tanveer, S. Virji, L. Andreeva, N. F. Totty, J. J. Hsuan, J. M. Ward, and M. Crompton (1996). 'Involvement of cyclophilin D in the activation of a mitochondrial pore by Ca2+ and oxidant stress.' *Eur J Biochem*, **238**(1), 166–172.

[176] S. Tharun, W. He, A. E. Mayes, P. Lennertz, J. D. Beggs, and R. Parker (2000). 'Yeast Sm-like proteins function in mRNA decapping and decay.' *Nature*, **404**(6777), 515–518.

[177] A. Thiele, S. Strkel, and H. E. Stierle (1998). '[Retroperitoneal fi brosis and arthritis–a manifestation of the same illness]'. *Z Rheumatol*, **57**(5), 285–291.

[178] N. Tjandra, J. G. Omichinski, A. M. Gronenborn, G. M. Clore, and A. Bax (1997). 'Use of dipolar 1H-15N and 1H-13C couplings in the structure determination of magnetically oriented macromolecules in solution.' *Nat Struct Biol*, **4**(9), 732–738.

[179] J. S. Toll (1956). 'Causality and the Dispersion Relation: Logical Foundations'. *Physical Review*, **104**, 1760 – 1770.

[180] K. Touhara, J. Inglese, J. A. Pitcher, G. Shaw, and R. J. Lefkowitz (1994). 'Binding of G protein beta gamma-subunits to pleckstrin homology domains.' *J Biol Chem*, **269**(14), 10217–10220.

[181] M. Tropschug, D. W. Nicholson, F. U. Hartl, H. Koehler, N. Pfanner, E. Wachter, and W. Neupert (1988). 'Cyclosporin A-binding protein (cyclophilin) of Neurospora crassa. One gene codes for both the cytosolic and mitochondrial forms.' *J Biol Chem*, **263**(28), 14433–14440.

[182] V. Tugarinov and L. E. Kay (2003). 'Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods.' *J Am Chem Soc*, **125**(45), 13868–13878.

[183] V. Tugarinov and L. E. Kay (2004). 'An isotope labeling strategy for methyl TROSY spectroscopy.' *J Biomol NMR*, **28**(2), 165–172.

[184] A. van den Boogaart, M. Ala-Korpela, J. Jokisaari, and J. R. Griffi ths (1994). 'Time and frequency domain analysis of NMR data compared: an application to 1D 1H spectra of lipoproteins.' *Magn Reson Med*, **31**(4), 347–358.

[185] L. Verlet (1967). 'Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules'. *Phys. Rev.*, **159**, 98 – 103.

[186] H. L. Vieira and G. Kroemer (1999). 'Pathophysiology of mitochondrial cell death control.' *Cell Mol Life Sci*, **56**(11-12), 971–976.

[187] T. Voets, G. Buyse, J. Tytgat, G. Droogmans, J. Eggermont, and B. Nilius (1996). 'The chloride current induced by expression of the protein pICln in Xenopus oocytes differs from the endogenous volume-sensitive chloride current.' *J Physiol*, **495 (Pt 2)**, 441–447.

[188] G. Vriend and C. Sander (1993). 'Quality Control of Protein Models: Directional Atomic Contact Analysis'. *J. Appl. Cryst.*, **26**, 47–60.

[189] W. Vuister, A. Wangand, and A. Bax (1993). 'Measurement of Three-Bond Nitrogen-Carbon J Couplings in Proteins Uniformly Enriched i 15N and 13C'. *J.Am. Chem. Soc.*, **115**, 5334–5335.

[190] P. Y. Wai and P. C. Kuo (2004). 'The role of Osteopontin in tumor metastasis.' *J Surg Res*, **121**(2), 228–241.

[191] S. Waldegger and T. J. Jentsch (2000). 'Functional and structural analysis of ClC-K chloride channels involved in renal disease.' *J Biol Chem*, **275**(32), 24527–24533.

[192] D. C. Wallace (1999). 'Mitochondrial diseases in man and mouse.' *Science*, **283**(5407), 1482–1488.

[193] A. M. Weljie, P. M. Hwang, and H. J. Vogel (2002). 'Solution structures of the cytoplasmic tail complex from platelet integrin ?IIb- and ?3-subunits'. *PNAS*, **99**, 5878–5883.

[194] D. S. Wishart, C. G. Bigam, A. Holm, R. S. Hodges, and B. D. Sykes (1995). '1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects.' *J Biomol NMR*, **5**(1), 67–81.

[195] M. Wittekind and L. Mueller (1993). 'HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha- and Beta-Carbon Resonances in Proteins'. *J. Magn. Res.*, **101**, 201–205.

[196] K. Wuethrich (1986). NMR of Proteins and Nucleic Acids. Wiley New York.

[197] G. M. Wulf, A. Ryo, G. G. Wulf, S. W. Lee, T. Niu, V. Petkova, and K. P. Lu (2001). 'Pin1 is overexpressed in breast cancer and cooperates with Ras signaling in increasing the transcriptional activity of c-Jun towards cyclin D1.' *EMBO J*, **20**(13), 3459–3472.

[198] T. Yamazaki, M. Yoshida, and K. Nagayama (1993). 'Complete assignments of magnetic resonances of ribonuclease H from Escherichia coli by double- and triple-resonance 2D and 3D NMR spectroscopies.' *Biochemistry*, **32**(21), 5656–5669.

[199] T. J. Ypma (1995). 'Newton-Raphson minimzation'. *SIAM Review*, **37**, 531?551.

[200] V. V. Zakharov, J.-P. Capony, J. Derancourt, E. S. Kropolova, V. A. Novitskaya, M. N. Bogdanova, and M. I. Mosevitsky (2003). 'Natural N-terminal fragments of brain abundant myristoylated protein BASP1.' *Biochim Biophys Acta*, **1622**(1), 14–19.

[201] N. Zamzami and G. Kroemer (2001). 'The mitochondrion in apoptosis: how Pandora's box opens.' *Nat Rev Mol Cell Biol*, **2**(1), 67–71.

[202] C. Zhang, J. P. Steiner, G. S. Hamilton, T. P. Hicks, and M. O. Poulter (2001). 'Regeneration of dopaminergic function in 6-hydroxydopamine-lesioned rats by neuroimmunophilin ligand treatment.' *J Neurosci*, **21**(15), RC156.

[203] Q. Zheng and Y. Zhao (2007). 'The diverse biofunctions of LIM domain proteins: determined by subcellular localization and protein-protein interaction.' *Biol Cell*, **99**(9), 489–502.

[204] M. Zoratti and I. Szab (1995). 'The mitochondrial permeability transition.' *Biochim Biophys Acta*, **1241**(2), 139–176.

[205] E. R. Zuiderweg and S. W. Fesik (1989). 'Heteronuclear three-dimensional NMR spectroscopy of the inflammatory protein C5a.' *Biochemistry*, **28**(6), 2387–2391.

Acknowledgement

At the end i pleased to special thank all persons which superivised me or contributed anyway else for the sucessful finalizing of this work.

- Prof. Dr. Robert Konrat for supervising, providing the position, and just for all.
- Prof. Dr. Georg Kontaxis for any kind of technical support and quick skillful help.
- To all my current and former colleagues, Dr. Elisabeth Dowler, Dr. Bettina Schweng, Dr. Paul Schanda, Dr. Agnes Koevari, Dr. RosaMaria Quintero, Dr. Matina Ortbauer, Dr. Karin Kloiber, Dr. Martin Tollinger, M.sc. Rebby Vie Colon and our Lab Manager Ing. Karin Ledolter for good working atmosphere.
- Mag. Erwin Gaubitzer, Mag. Bashir Khan, Mag. Bettina Spitzenberger, Mag. Anita Salmazo, Dr. Patrizia Abrusci, Mag. Mads Beich-Frandsen, Ing. Claudia Schreiner, Dr. Bjoern Sjoeblom for any kind of useful discussion.
- Ing. Werner Koenig for quick assisting concerning hard ware problems.
- Dr. Rosaria Gandini for excellent preparation of ICln samples used in this work.
- Dr. Hannes Fuerst for competent and kindly cooperation work and help support regarding biological aspects of ICln.
- Austrian WWF, Proceryon l.l.c, and Vienna WWTF for funding.